34368

National Library of Canada | Bibliothèque nationale du Canada | **CANADIAN THESES ON MICROFICHE** | ***THÈSES CANADIENNES SUR MICROFICHE***

NAME OF AUTHOR/*NOM DE L'AUTEUR* KEITH DOUGLAS HORTON

TITLE OF THESIS/*TITRE DE LA THÈSE* DEPTH OF PROCESSING: THEORY AND RESEARCH

UNIVERSITY/*UNIVERSITÉ* OF ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* PhD

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1977

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* DR W.N. RUNQUIST

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* AUG 3/77 SIGNED/*SIGNÉ*

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ*
DEPT OF PSYCHOLOGY
WILFRID LAURIER UNIVERSITY
WATERLOO, ONT N2L 3C5

NL-91 (3-74)

National Library of Canada
Cataloguing Branch
Canadian Theses Division
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

NL-339 (3/77)

THE UNIVERSITY OF ALBERTA

DEPTH OF PROCESSING:
THEORY AND RESEARCH

by

©

KEITH DOUGLAS HORTON

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY
EDMONTON, ALBERTA

FALL, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Depth of Processing: Theory and Research" submitted by Keith Douglas Horton in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Willie Runquist
Supervisor

....

....

R.B. Wilberg

....
External Examiner

DATE....July 27, 1977....

ABSTRACT

The theoretical development of the recent depth of processing model is reviewed in detail, along with evidence pertaining to many aspects of the model. Comments are addressed primarily to features of the theory which as yet are insufficiently developed to permit unequivocable empirical predictions and tests. The research was directed to a number of theoretical matters for which there was little or no evidence. These included the role of reconstruction in long-term retrieval, the significance of a minimal core encoding for memory, and the memorial effects of experience with the depth of processing research paradigm. Very generally, evidence was found in support of reconstruction, although other processes also seemed necessary to account for all retrieval. Secondly, in line with much data, there was little evidence of a memorial representation resulting from the core encoding. Finally, it appeared that subjects will learn to cope with the encoding task (in terms of later recall performance) by performing additional semantic analyses on items if the test conditions permit its use.

ACKNOWLEDGMENTS

A number of people played significant roles in the various stages of this research program. Both Drs. Peggy Runquist and Steve Carey helped to improve the design of the experiments and commented on an earlier draft of this paper. My supervisor, Dr. Willie Runquist, was a particularly competent and scholarly source of advice throughout all stages of the program, not to speak of his financial support for the entire research endeavour.

Experiments 1, 2, and 5 were run by Donna Ethier and Experiments 3 and 4 by Sandy MacPherson. Both were instrumental to the smooth execution of the research.

My wife's role in this dissertation (indeed, in my entire graduate career) makes itself evident in many different forms. Tangibly, there were the hundreds of computer cards punched; there were the hours spent diligently scoring and re-scoring page upon page of data sheets; there were the late nights and long weekends unselfishly passed at her second home, in front of the computer terminal. These along with many other intangible contributions cannot pass unacknowledged.

TABLE OF CONTENTS

PAGE

INTRODUCTION 1

Criticisms of the Multi-Store Model 3

Depth of Processing 7

Evidence. 15

Comments. 33

Revisions to the Theory 41

Evidence. 70

Comments. 81

Overview of the Research. 89

EXPERIMENT 1

Introduction. 92

General Method. 93

Method. 97

Results 99

Discussion. 110

EXPERIMENT 2

Introduction. 118

Method. 122

Results 125

Discussion. 132

EXPERIMENT 3

Introduction. 134

Method. 137

Results 139

Discussion. 141

EXPERIMENT 4

Introduction. 144

Method. 146

Results 147

Discussion. 151

EXPERIMENT 5

Introduction. 154

Method. 157

Results 160

Discussion. 169

GENERAL DISCUSSION 172

REFERENCES 191

APPENDIX I. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 1: FREE RECALL 205

APPENDIX II. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 1: RECOGNITION (HITS). 206

APPENDIX III. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 2: FREE AND CUED RECALL. 207

APPENDIX IV. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 2: FREE RECALL 208

APPENDIX V. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 3: CUED RECALL (ALL GROUPS). 209

APPENDIX VI. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 4: RECOGNITION 210

APPENDIX VII. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 5: IMMEDIATE RECALL ACROSS TRIALS IN THE 3-TRIAL GROUPS. 211

APPENDIX VIII. ANALYSIS OF VARIANCE SUMMARY TABLE EXPERIMENT 5: FINAL CUED RECALL WITH SEMANTIC CUES. 212

LIST OF TABLES

TABLE	DESCRIPTION	PAGE
1	Experiment 1: Percent Free Recall as a Function of Instructional Condition, Encoding Condition, and Type of Encoding Information Recalled	178
2	Experiment 1: Percent Correct Recognition as a Function of Instructional Condition, Encoding Condition, and Type of Encoding Information Recalled.	179
3	Experiment 1: Mean Confidence Judgments for Hits and False Alarms as a Function of Encoding Condition. Data are Averaged across Instruction Conditions.	180
4	Experiment 2: Percent Recall as a Function of Encoding and Test Condition (Groups 1 and 2). . . .	181
5	Experiment 2: Percent Recall as a Function of Study and Test Condition (Congruent Encodings Only) . . .	182
6	Experiment 3: Examples of Test Cues used by the Three Groups of Subjects in Each of Three Encoding Conditions.	183
7	Experiment 4: Mean Confidence Judgments (number of observations included in brackets).	184
8	Experiment 5: Percent Recall for the 3-trial Groups as a Function of Encoding and Test Conditions . . .	185

LIST OF FIGURES

FIGURE DESCRIPTION PAGE

1 Experiment 2: Recall of target words as a function of study condition and level of encoding. 186

2 Experiment 2: Retention of encoding information by the free recall groups as a function of study condition, level of encoding, and type of target information recalled. 187

3 Experiment 3: Cued recall as a function of level of encoding and test condition 188

4 Experiment 5: Final cued recall with phonemic cues as a function of level of encoding, congruity, number of trials, and immediate test condition. . . 189

5 Experiment 5: Final cued recall with semantic cues as a function of level of encoding, congruity, number of trials, and immediate test condition. . . 190

Throughout the 1960's, memory research became increasingly dominated by an information-processing view of the perceptual-memorial system. The conjoint insurgence of this view with the advent of the computer age was undoubtedly more than just coincidental. The essential characteristic of this information-processing approach and computer systems is the concept of "information flow": input progesses through the various points, or stages, within the system. At the various stages, some processing of the information occurs, after which it advances to the next stage.

Within this paradigm (cf. Kuhn, 1962) for memory theory and research, the most prevalent approach was what has variously been termed the structural (of temporally-structured) models (e.g., Murdock, 1974), modal models (Murdock, 1967), or multi-store models. There are innumerable examples of these theories but probably the most influential have been those of Broadbent (1958), Waugh & Norman (1965), and Atkinson & Shiffrin (1968, 1971). Basic to this general class of theories is the idea of memory stores, or repositories. Typically, three such stores are identified: sensory memory, short-term memory (STM), and long-term memory (LTM). The evidence for each of these stores has been reviewed in numerous sources and need not be repeated here. However, a brief description of some of the characteristics of each store will be presented. Unless otherwise stated, each characteristic may be regarded as generally true for most multi-store models. We will consider here only those characteristics which have been dealt with in recent criticisms of the theory and which will be of particular interest in terms of the primary theoretical concern of this paper, the depth of processing model of memory. Consideration of this more contemporary approach to memory theory will be delayed until the theoretical and empirical background for it has been reviewed.

Information enters the system through the sensory store. The sensory store is considered to be a large capacity memory from which information is lost through decay at a very rapid rate (usually one second or less) unless it is "attended". The relatively raw perceptual format of coding typically allocated to sensory memory is reflected in the use of the terms "iconic memory" and "echoic memory" for the visual and auditory sensory memories respectively.

Attending to information in sensory memory results in it being entered into STM. The capacity of STM is said to be not less than three items and probably not more than seven items. Information will be lost from STM over a period of 5-10 seconds unless it is "rehearsed". As long as rehearsal is occurring, the onset of information loss can be delayed indefinitely. With regard to the format of coding in STM, there is as yet little or no agreement. Some authors have argued that items are retained in STM solely on the basis of their phonemic codes (e.g., Baddeley, 1972; Conrad, 1967; Glanzer, 1972) while others have not restricted the format of coding in this way (e.g., Atkinson & Shiffrin, 1971; Shulman, 1971; Waugh & Norman 1965). While attention is the process or mechanism by which information is transferred from sensory memory to STM, simple rehearsal frequency is often viewed as the necessary and sufficient condition for the transfer of information from STM to LTM.

Once information has been encoded into LTM, it is usually considered to be there permanently: forgetting becomes a matter of failing to retrieve a target which actually exists in memory rather than the memory trace being lost or destroyed in some way. These two sources of "forgetting" have been labelled as problems of accessibility and availability respectively (cf. Tulving & Pearlstone, 1966). Long-term memory has an unlimited capacity for storing information. The format of coding is sometimes considered to be strictly semantic (e.g., Baddeley, 1972) however it has recently been argued that all forms of coding exist in LTM (Horton, 1976; Shulman, 1971).

Criticisms of the Multi-Store Model

The criticisms of this view of the human memory system have taken a variety of different approaches. Bernbach (1975) and Wickelgren (1973) have both elaborated on theoretical problems with the multi-store model. Wickelgren has argued that much of the evidence that has been used to distinguish two stores is at best equivocal (e.g., the issue of a phonemic STM and a semantic LTM[1]) or could readily be interpreted in terms of a more parsimonious one-process theory (e.g., the serial position effect in free recall). This latter issue had previously been spoken to by Melton (1963). Bernbach (1975) has pointed out that two incompatible views of the STM system have been posited and that one can derive evidence in support of either under certain conditions. He was led to the conclusion that this is a problem for all multi-store theories and implies a preference for a single process theory.

Other authors have argued against the multi-store approach on empirical grounds. For example, Shulman (1971, 1972) has found that, when the nature of the task demands it, semantic information can be found in what are sometimes considered to be STM testing conditions.

In a more comprehensive empirical critique of multi-store theory, Craik (1973; Craik & Lockhart, 1972) has noted a number of potential problems involving the issues of capacity, coding, transfer, and forgetting characteristics. With regard to capacity, Craik points out that estimates of STM capacity have ranged from 2 to 20 items and that "if capacity is a critical feature of STM operation, a box model has to account for this very wide range of capacity estimates" (Craik & Lockhart, 1972, p. 673). On the issue of coding, a number of papers were cited each suggesting one or more of a number of possible codes for STM storage. These include acoustic, semantic, and visual.

---

[1] Posner & Warren (1972) have also argued against the logic of the distinction if phonemic and semantic coding is to be the discriminating feature.

There has been, to date, a substantial amount of evidence directly contradicting the proposal that probability of storage in LTM is a direct function of amount of rehearsal of information in STM (e.g., Jacoby, 1973; Craik & Watkins, 1973). Consequently, Craik suggests that this basis for transfer from STM to LTM is not supported by recent investigations. Finally, with respect to forgetting characteristics, Craik and Lockhart cite a variety of studies which show a wide range of forgetting functions for both STM and sensory memory. They argue that the retention functions should be relatively invariant across conditions if the stores are to be viable and distinguishable.

While there are many additional criticisms of multi-store theory in the literature, the brief selection above might be classified as representative if not extensive. Since our present interest is with the development of the depth of processing model and, therefore, those aspects of the multi-store which Craik suggests are problematic, a few brief comments on the validity of his criticisms are in order.

First, it should be noted that many new theories have recently appeared in the literature which could be classified as updated versions of the multi-store theory. These theories envelop the concepts of STM and LTM although sometimes using different labels such as "working memory" and "semantic memory" respectively. An interesting aspect of some of these theories is that STM is often viewed as simply that part of LTM which is currently activated (e.g., Anderson & Bower, 1973; Kintsch, 1972; Norman, 1968). Positing such a relationship between STM and LTM eliminates the necessity for the concept of transfer from one store to the other. At a more general theoretical level, these authors have attempted to minimize the concept of "memory stores" as static repositories and rather emphasized the processes occurring within them. Hence, Anderson and Bower (1973) and Weist (1972) use the term "working memory" to refer to STM.

In response to a number of empirical and theoretical criticisms, these authors have not identified phonemic and semantic information with STM and LTM respectively. In fairness, it should be noted that this was not, in fact, a tenet of all of the more traditional multi-store models, but only a few. Thus, such a criticism has limited applicability.

Finally, there is the problem of ascertaining which items are retrieved from which store at the time of testing. Before arguments over capacity, forgetting characteristics, and even the format of coding can be resolved, some general agreement must be reached on an a priori basis for distinguishing items retrieved from STM and LTM. While a number of alternatives have been offerred (see Watkins, 1974a, for a review), there has been no concensus.

Criticisms of the multi-store models which necessitate such a measure serve to highlight the incomplete theoretical development of these models rather than to debase them empirically.

Depth of Processing

In proposing their alternative to the multi-store models, Craik and Lockhart (1972) consider first the nature of perceptual processing and the establishment of the memory trace. When a stimulus enters the system, it undergoes "a series or hierarchy of processing stages" progressing from early stages involving "such physical or sensory features as lines, angles, brightness, pitch, and loudness" through to so-called deeper analyses "concerned with pattern recognition and the extraction of meaning" (1972, p. 675). These different levels of analysis are seen as existing on a continuum (Craik, 1973, p. 49). There appears to be a certain rigidity in this sequence of perceptual analyses such that each level of analysis must occur in sequence.

Once the perceptual analysis has resulted in stimulus "recognition" (by this, they seem to mean conscious awareness), additional processing may occur. These Craik and Lockhart identify as the "semantic-associative stages of stimulus enrichment" (p. 675). They include the arousal of associations and images. Unlike those analyses occurring prior to recognition, these later "post-perceptual" analyses do not "occur as a sequence of necessary steps, but rather as one or several of many possible options" (Craik, 1973, p. 50). The option selected is under the subject's control. To emphasize the flexibility of these later analyses, Craik suggests that "spread" might be a better term than "depth" since it implies that processing may proceed in any of a variety of directions. However, he retains "depth" for other reasons.

This general view of how an item is processed is not peculiar to depth of processing. Indeed, many authors have adopted such a system as we will note later.

As a direct result of the processing enacted on an item, a memory trace is established. The products of each stage in the analysis automatically become a part of the trace for that item. There is clearly an emphasis here on the continuity between perceptual processes and memory traces. This close interrelationship has been evidenced in a number of other sources. Broadbent (1958) was one of the earliest of modern theorists to emphasize the bond between perception and memory. Other workers in the field of selective attention have followed his lead (e.g., Norman, 1969; Treisman, 1964). More recently, Jenkins (1974a; 1974b) has addressed himself to this issue directly. He has also suggested that perception and memory are inseparable. However, he takes the matter even further by positing that both may be inseparable from some of the higher cognitive processes such as imagining and problem solving. Others have more or less implicitly assumed the significance of the interrelationship. This is particularly true of the many models dealing with the variability of encoding of items (e.g., Bower, 1967; Martin, 1968). Wicken's (1970, 1972) work can also be seen as following from this assumption.

Information extracted from the various levels of analysis is retained differentially: the products of preliminary analyses are lost rather quickly but deeper level information is lost at a much slower rate. Thus the most effective analysis in terms of an item's later memorability is that which occurs post-perceptually in the subject's semantic-associative system.

Implict in the early statements of the theory was the view that the subject could control the depth to which an item was processed. Thus, if the subject was asked for some structural information about an item, he could terminate the processing at that point and thereby never reach deeper semantic levels of analysis for that item. The closest Craik comes to acknowledging this cognitive control of depth of coding is in reference to an experiment using the depth of processing paradigm in which subjects answer questions about particular features of words presented to them. He notes that "the different questions would necessitate processing to a progressively deeper level" (Craik, 1973, p. 57). Thus it would appear that the depth to which an item is processed could be classified as a control process (Atkinson & Shiffrin, 1968) in this version of the theory, although it is not explicitly identified as such by the authors. Since subjects are not usually aware of the impending recall test using this procedure, there would be no need for them to elaborate the processing of an item beyond that necessary to answer the question. However, in a typical free recall experiment in which subjects are explicitly told to learn the words for the test trial, it would be expected that they would process the words to a much deeper level so as to maximize performance. Thus, "the instruction to learn facilitates performance only insofar as it leads the subject to process the material in a manner which is more effective than the processing induced by the orienting task in the incidental condition" (Craik & Lockhart, 1972, p. 677).

The essence of the depth of processing approach as it was first presented in 1972 revolved around the nature of information processing and its relationship to memorability. However, a related issue discussed there dealt with the nature of rehearsal and the role it plays in retention. The authors identified two forms of rehearsal which they simply referred to as Type I and Type II. Type I rehearsal, or processing, refers to the "repetition of analyses which have already been carried out." That is, the subject is simply maintaining an item at a given level within the system by continuously reanalyzing it at that level. It is referred to also as maintenance rehearsal since its principal function is to hold an item at a high level of accessibility over the short term. There is an obvious continuity with more traditional views of rehearsal: 'such descriptions as "continued attention to certain aspects of the stimulus," "keeping the items in consciousness," "holding the items in the rehearsal buffer," and "retention of the items in primary memory" all refer to the same concept of maintaining information at one level of processing' (Craik & Lockhart, 1972, p. 676).

Items which are being maintained in consciousness are referred to as primary memory (PM) items although PM is not a memory store as in the earlier models but rather a mode of storage. Note that the nature of the memory codes of items in PM will be strictly a function of the level at which the processor is operating. Thus, if items are being rehearsed at a phonemic or semantic level, the memory codes will be phonemic or semantic respectively. Further, they suggest that certain codes (e.g., phonemic) are easier to maintain than other codes (e.g., early structural analyses) and, since the processor has a limited capacity with which to work, the amount of information in PM at any one time will be a function of the type of information being held there. Viewing PM in this way permits Craik and Lockhart to deal with the two issues of coding and capacity which they saw as problematic for multi-store theories.

Since Type I rehearsal involves maintaining items at a particular level within the system, it does not result in an enriched and more durable trace. In contrast to this, Type II rehearsal involves the execution of additional deeper level analyses which will lead to an increasingly durable trace. The definition of Type II processing is left somewhat less concrete than that of Type I but basically it appears to be a relative matter: "Type II processing... involves deeper analysis of the stimulus" (Craik & Lockhart, 1972, p. 676); "Type II processing...involves further processing of the stimulus to a deeper level" (Craik, 1973, p. 51). Thus, whether a subject is engaging in Type I or Type II rehearsal at any given point in time will be determined strictly by the nature of the ongoing processing: if he is reanalyzing a stimulus for information which is already available, then Type I processing is implicated; if he is analyzing for new information, then it is Type II processing. The operational evidence for this distinction will be cited later.

It should be evident from this discussion that Type I and Type II processing are represented as control processes in the system as was the case with the depth of processing of an item. These two control processes are clearly interrelated to the extent that the depth to which an item is processed will be a function of the type of rehearsal, Type I or Type II, which is initiated.

Other authors have also dealt with the distinction between these two types of rehearsal, although using different terminology. Tulving and Madigan (1970) identified "elaboration coding" which they define as "the storage of additional nonredundant information with the verbal unit" and accorded it the same memorial function as

Type II rehearsal.[2]

Rumelhart, Lindsay, and Norman (1972), without using a particular label, made the distinction with reference to the Mandler and Dean (1969) "one plus one" experimental procedure. The following quotation illustrates their use of two types of rehearsal or processing:

> The general strategy for learning was to start off trying to remember the specific items on the list in STM and to reorganize the STM information whenever STM capacity was exceeded. Reorganization was based on examining the classes to which current residents of STM belonged, attempting to replace specific items with their class names whenever this would reduce the amount of information that had to be remembered. (p. 238)

Thus, higher order reorganization on the basis of semantic classes is an example of Type II processing whereas the simple maintenance of specific items in STM represents Type I processing. Clearly Type II processing is seen as a control process by Rumelhart et al., as it is with Craik and Lockhart.

Weist (1972) and Kintsch (1972, 1974) both take a somewhat different approach, although maintaining the distinction. For example, Weist states that "rehearsal provides the occasion and the mechanism for changing as well as maintaining working memory representations, and revising the familiarity value of the features of associated lexical concepts" (p. 441).[3] The "maintaining" and "changing" functions of rehearsal correspond to Craik's use of Type I and Type II rehearsal respectively. Type II rehearsal, by which Weist means reorganization and additional semantic analyses specifically, is viewed as a more or less automatic result of similarity among items in working memory, or STM. Thus, similarity is a sufficient condition for Type II rehearsal. Kintsch takes a very similar view in terms of the automatic nature of Type II processing. Clearly, both authors are in disagreement with Craik with regards to how elaborative processing is invoked, but all concur with the conceptual distinction between Type I and Type II processing and the subsequent effect each has on memorability. It is clear, then, that the idea of different types of rehearsal and their respective effects on retention curves is not unique to depth of processing. Indeed, a substantial amount of agreement exists among many authors with respect to this issue.

---

[2] Actually, Tulving and Madigan distinguish "substitution coding" and "elaboration coding", although these terms do not correspond with Type I and Type II processing. Indeed, from their description, substitution coding could largely be classified as Type II processing by Craik's use of the term. These two forms of coding identified by Tulving and Madigan are to be contrasted with simple rehearsal to which they later refer when discussing the Atkinson and Shiffrin (1968) theory.

In summary, we would simply repeat what appear to be the three significant characteristics of the depth of processing model as it was originally elaborated. First, information entering the system is processed through a series of analyses beginning with the extraction of gross physical information and concluding with relatively elaborate semantic codes. The subject appears to be able to control the depth of processing carried out on an individual item. Stimulus recognition is the result of this processing. Secondly, the depth to which an item is processed will determine the long-term retention characteristics of that item: the greater "depth" an item reaches in the system, the better its memorability. Finally, two types of rehearsal, Type I and Type II are distinguished. According to Craik, Type I involves merely reanalyzing an item for information which has already been extracted and results in high accessibility of the item over the short term but no long term retention increment. In contrast, Type II processing denotes additional analyzing of information in which the subject may choose to engage. This processing appears to be restricted to the "semantic-associative" system. Items which undergo Type II processing will show improved long-term memorability. Short-term retention may be reduced somewhat due to a reduced capacity of the central processor to retain items while devoting much of its capacity to deeper level processing.

---

[3] A "lexical concept" is merely the constellation of attributes possessed by a given item.

Evidence

A multitude of evidence exists dealing with the various features of the depth of processing model. To document it all would be an extraordinary task. Rather, we will discuss only some of the most salient findings which focus primarily on two aspects of the theory.

The first issue we will discuss is the correlation between depth of processing and memorability. The prime evidence cited in support of this hypothesized correlation derives from what we will heretofore call the depth of processing paradigm. The procedure is quite simple. Subjects are asked a question about a word which appears briefly. Often, subjects are told only that the experiment is dealing with the perception of word attributes and no memory test is mentioned. Thus, learning is often incidental. A variety of different types of questions are typically asked. These include questions dealing with structural information about the word (e.g., Does it contain the letter _____?; Is it written in uppercase?; Does it begin with the same letter as _____?), phonemic (Does it rhyme with _____?), semantic (Is it a member of the category _____?; Where does it stand on a 1-7 scale of pleasantness?), or deeper level sentential information (Does the word fit into the following sentence: __________?). It is presumed that each of these types of questions deal with successively deeper levels of information.

Numerous studies using essentially this procedure have found that depth of processing is in fact highly correlated with memorability (e.g., Craik & Tulving, 1975; Gardiner, 1974; Mondani, Pellegrino, & Battig, 1973; Seamon & Murray, 1976; Till, Cormak, & Prince, 1977; Walsh & Jenkins, 1973). In addition, a number of authors have employed major variations on this procedure and still obtained the predicted differences in performance (e.g., Bellezza, Richards, & Geiselman, 1976; Cermak & Youtz, 1976; Nelson, Wheeler, Borden, & Brooks, 1974). Indeed, there seem to be only a few situations in which the phenomenon has not been evidenced. The first deals with Jacoby's (1974) "looking-back" procedure in which subjects attempt to identify recently presented words having a particular feature in common with the current word. Using comparable experimental procedures, Bregman (1968) and Shulman (1970) also failed to observe retention effects due to depth of processing, although other research of this type has in fact found the effect (Cormak & Youtz, 1976). Because of their relevance to later theoretical developments, we will discuss these below.

Secondly is the finding by Wetzel (1975) that depth of processing did not have the predicted effect in the directed forgetting paradigm (cf. Bjork, 1972). In a between-subjects design, Wetzel required his subjects to simply rehearse to-be-remembered (TBR) items, generate a rhyme (Experiment 1 only), or generate an adjective or noun (semantic task) associated with the presented item. Both in immediate and final free recall, the rehearsal and rhyme groups performed better than the semantic group on recall of the TBR's. There was no significant or consistent trend with the to-be-forgetten (TBF) items, probably due in part to the very low performance level in both immediate and final recall tests.

This finding was at odds with so much reported data dealing with depth of processing that it was felt necessary to attempt a replication in our own lab. We felt that perhaps there was something sufficiently different about the directed forgetting paradigm to offset the effect of depth and it would be interesting to isolate this factor. First, however, a replication was in order. Using a design basically identical to Wetzel's but with some additional variables, Horton & Petruk (unpublished) obtained contradictory findings. When subjects were required to perform a structural or phonemic task, recall levels were substantially reduced compared to a semantic task. The minimal procedural differences in the experiments by Wetzel and our own do not seem to provide a clue as to why the different findings obtained. Therefore, at this point, it seems wise to withold judgment on the effects of depth of processing in the directed forgetting paradigm.

The final set of data which may or may not be relevant to the depth-memorability issue deal with items to which minimal "attention" is accorded. As an example, Moray (1959) found that items presented repeatedly on an unattended channel in an auditory shadowing task were not recognized above chance upon completion of the task. Norman (1969b) found that items from an unattended channel were "available" for a few seconds after presentation but were lost quickly after that (see also Moray, 1959). Since it seems that information on the unattended channel may be processed semantically (Treisman, 1960; 1964), it is unclear as to why Moray and Norman did not observe any retention if in fact semantic analysis results in a relatively long lasting, durable memory trace. Wickens, Moody, and Shearer (1976) have very recently replicated these findings by showing that even though semantic information is extracted from the unattended channel (Experiment 2), it is not remembered (Experiments 1 & 2) even when items are repeated (Experiment 3).

In a similar vein, Neisser and Beller (1965) presented subjects lists of words from which they were to locate a single target meeting some predefined semantic criteria. They also observed essentially no retention of rejected items, despite the fact that these items must be coded to some minimal semantic level in order to be rejected. This led the authors to the conclusion that rejected items were not even stored, a view perhaps different from that of Craik who sees perception of an item as resulting in automatic storage. Schulman (1971) suggests that the low retention levels for rejected items might be due partly to the processing of the list items in general. Subjects failed to detect the target 17% of the time. Thus Schulman concludes that "such a high percentage of misses may reflect gaps in the scan that inevitably would result in the non-recognition of unscanned words." Using a task in which subjects overtly acknowledged each item as a target or non-target (half the items presented were targets), Schulman reduced error rates during scanning to about 3% or less. Concurrently, forced-choice recognition of non-targets, even after a structural encoding task, was well above chance levels. For the semantic encoding task, non-target recognition was more than 50%, corrected for guessing, which clearly implicates storage of these items.

Since we cannot be sure as to how subjects are processing information in these paradigms, we hesitate to suggest that it is directly applicable to Craik's formulations.

Tangentially associated with this issue of depth of coding and memorability is the relationship between the time required to answer the encoding question (which will be referred to simply as the reaction time, or RT) and depth of coding. Jenkins has put the issue very neatly:

> We had a good deal of feedback from colleagues who offered two common explanations for our findings. One explanation was that the semantic orienting tasks were easier than the nonsemantic orienting tasks. It was argued that this difference permitted the semantic subjects more time to rehearse or think about the words. The other explanation was just the opposite: The semantic orienting tasks required more effort, and thus the subjects worked on the words more, attended more, and processed more. All this effort, then, resulted in better recall. The first hypothesis we can call the time hypothesis, and the second hypothesis we can term the effort hypothesis. (Jenkins, 1974b, p. 12; emphasis is original)

Walsh and Jenkins (1973) offered one empirical solution to these conflicting hypotheses, however it seems reasonable to deal with these issues strictly in terms of time to perform the respective orienting tasks. Given that the time hypothesis predicts smaller RT's and the effort hypothesis predicts greater RT's to perform the semantic tasks than the lower level tasks if the depth effect is to occur, evidence can be provided to discredit both alternative interpretations. For example, Craik (1973, Experiments 4 & 5; Craik & Tulving, 1975, Experiments 1 & 2) has observed the depth effect when RT was directly related to the depth of processing necessitated by the type of orienting task. These data bear directly on the time hypothesis which leads to the opposite prediction. Gardiner (1974) required subjects to perform either a phonemic or semantic orienting task. He found that even though RTs were substantially longer for the phonemic task, free recall performance was still more than double for items encoded with the semantic task. Mondani et al. (1973) and Schulman (1971) obtained similar results when they compared structural and semantic orienting tasks. These data are clearly in contrast to that predicted by the effort hypothesis.

Craik and Lockhart addressed themselves briefly to the relationship between RT and depth of coding. They suggested that the depth to which an item is processed will normally be correlated directly with RT's. It seems however, on the basis of the data described above, that lower level tasks can be devised which require more or less processing time relative to a semantic task. Despite this, the evidence clearly shows that it is in fact depth and not processing time which is critical to memorability. Whether this will still be the case with extremes in processing time is uncertain. However, within the limits set by current research, the depth effect can truly be attributed to the nature of the task rather than the time to perform it.

A final related issue is the hypothesized continuity between perception and memory. Probably the best data which can be cited in support of this view is the research using incidental learning procedures. Specifically, we are interested in what Postman (1964) has identified as Type I incidental learning, not to be confused with Type I rehearsal or processing. The essential characteristic of this procedure is simply that subjects are exposed to the stimulus materials but are not given any instructions about learning them. Rather, the experimenter requires them to perform some orienting task on the items. Type I incidental learning is to be contrasted with Type II in which subjects are instructed to learn some aspect of the materials presented and are then tested on some other, irrelevant, aspect.

Postman has summarized most of the literature on Type I incidental learning until 1964. To illustrate the potential effectiveness of various orienting tasks on incidental learning, he posits that "one may conceive of a continuum of orienting tasks, ranging from those requiring responses maximally favorable to learning to those requiring responses maximally antagonistic to learning" (Postman, 1964, p. 188). Craik and Lockhart (1972) in their early version of depth of processing would likely not disagree with this basic statement, although they might choose to substitute "retention" for learning. As they see it, the various levels to which an item could be processed was viewed as a continuum (Craik & Lockhart, 1972, p. 676) with the effectiveness of each level, in terms of memorability, being merely quantitative in nature. Thus, a structural orienting task would prove "maximally antagonistic" to retention whereas a semantic task would prove "maximally favorable".

In comparing incidental and intentional learning conditions, Postman concludes that:

> Intent per se has no significant effects on learning. All its effects are indirect, i.e., instructions to learn activate responses to the materials which are favorable to acquisition. The same results can be achieved by appropriate orienting tasks without instructions to learn. (Postman, 1964, p. 189; emphasis is original)

Before noting recent research on this matter, it should be noted that few researchers have followed Postman's early dictum in which he carefully pointed out that an appropriate analysis of the effect of a given orienting task can only be accomplished if both incidental and intentional learners perform it. Then a comparison may be made with an intentional learning group which does not perform the orienting task. This methodological consideration proves itself non-trivial in recent work. Johnston and Jenkins (1971), for example, found performance differences between intentional learning groups which did and did not perform a semantic orienting task requiring the generation of an adjective associated with the presented noun, or vice versa. Since this comparison becomes important in the later development of the depth of processing model and the concept of reconstruction current researchers would be well-advised to heed Postman's comments.

Within the framework of depth of processing, then, the appropriate comparison is between groups performing the same orienting task under incidental and intentional learning instructions respectively. It is presumed that intentional learning groups would carry on processing the information even after the requirements of the orienting task have been met. Any retention differences would thus be a result of the additional Type II processing in which the intentional groups engage themselves according to depth of processing.

A number of studies have observed incidental/intentional retention differences in free recall following a structural (Johnston & Jenkins, 1971; Mondani et al., 1973; Wolk, 1974) or a phonemic (Jacoby & Goolkasian, 1973) orienting task. These differences are also observed in recognition following a structural task (Wolk, 1974). Following different types of semantic orienting tasks, however, retention as measured by free recall is not dependent on whether learning instructions are given (Jacoby & Goolkasian, 1973; Johnston & Jenkins, 1971; Mondani et al., 1973). These findings clearly support the depth of processing position that incidental learning instructions can be as effective for information storage as intentional instructions if the processing required by the orienting task is at a level comparable to that achieved by the latter instructions. Differences in performance between intentional learners who do and do not perform the orienting task (Johnston & Jenkins, 1971) may be attributed to the amount of deeper level information extracted by the two groups. Thus, for a constant presentation interval, those subjects not performing the orienting task may have a greater opportunity to impose various mnemonic strategies (cf. Morris & Stevens, 1974) compared to subjects who must spend a significant part of their processing time analyzing the presented item for the information required by the orienting task.

It is interesting to note that Warrington and Ackroyd (1975) actually found superior retention (as measured by a yes/no recognition test) of words and faces by an intentional learning group making pleasantness ratings compared to an intentional learning group performing no orienting task. Further, an intentional learning group performing a structural orienting task* recognized the stimuli slightly, but not significantly, better than the group which did not perform the orienting task. It appears that this may be the only published study in which this relative standing of groups was observed. As such, the findings should be considered cautiously, although Craik and Lockhart allow for this possibility: "with an appropriate orienting task and an inappropriate intentional strategy, learning under incidental conditions could be superior to that under intentional conditions" (1972, p. 677). In addition, the unusual subject pool (aged 50-70 years), experimental setting (a crowded out-patient room in a hospital), and relatively poor recognition by the control group (75% over a 50-word list) should all be noted in considering the generality of these findings and whether they could be replicated.

---

*Whether these tasks are truly structural is uncertain. Categorizing black words on white cards as green or red and judging faces as tall or short raise some doubt as to exactly where to place these groups in terms of a continuum of orienting tasks.

With respect to the continuity between perception and memory, this research has shown that merely processing an item for certain types of information leads directly to some retention for that item. The amount of retention varies as a direct function of the level to which the item is analyzed, in accord with the depth-memorability correlation. Thus, as Postman suggests, intent to learn is not a necessary condition for retention.

The second major aspect of the theory for which we will describe empirical support is the distinction between Type I and Type II rehearsal. The point of making the distinction was that, in Craik's view, rehearsal may or may not lead to increased retention depending on the nature of the rehearsal. This is in contrast to the multi-store theorists who generally suggested that any rehearsal would automatically lead to increased transfer of information from STM to LTM.[5] The evidence for the distinction can be found in reports of two highly related phenomena, the negative recency effect and the storage-coding tradeoff. The negative recency effect refers to a phenomenon often observed in a test of final free recall (FFR) subsequent to immediate testing on a series of single trial free recall lists. Whereas the typical recency effect obtains on the immediate recall trials, performance on these same items in the FFR test may be depressed relative to the asymptotic level of the middle items when collapsed across all lists presented. Negative recency refers to this depressed recall for the recency items in each list. A number of researchers have reported this effect (Craik, 1970; Jacoby & Bartz, 1972; Maskarinec & Brown, 1974; Mazuryk, 1974; Roenker, 1974). Furthermore, requiring subjects to engage in additional Type I rehearsal of the recency items prior to the immediate free recall (IFR) test usually does not result in increased retention of these items, contrary to many multi-store theories (Jacoby, 1973; Mazuryk, 1974; Meunier, Kestner, Meunier, & Ritz, 1974; Roenker, 1974). Alternatively, if subjects are instructed to engage in Type II rehearsal of the last few items for a short interval prior to testing, a clear postive recency effect is observed even in FFR (Modigliani & Seamon, 1974; Mazuryk, 1974). These latter effects are not unexpected on the basis of multi-store theories since presumably this Type II rehearsal involves the activation of information stored in LTM.

---

[5] Actually, this hypothesized relationship between rehearsal and long-term retention is not restricted to multi-store models. An excellent example of this as a rather general belief is Underwood's statement of frequency theory (e.g., Ekstrand, Wallace, & Underwood, 1966) in which, again, it is suggested that simple repetition of items leads to increased memorability.

In preparing for the IFR task, it has been suggested that subjects quickly learn to maximize their output by recalling the last few list items first and then returning to earlier items. This permits subjects to minimize the amount of processing capacity allotted to recency items since in contrast to earlier items they will be able to survive the very short retention interval with relatively low level phonemic codes. Given that this type of code is not retained particularly well over the long term, negative recency effects in FFR might be predicted. On the other hand, pre-recency items must be coded more durably during the study interval and therefore are recalled somewhat better on the FFR test. Thus, requiring subjects to simply recycle recency items at a low level within the processing system would not be expected, according to depth of processing, to yield increased retention in the FFR test whereas coding these items at a deep semantic level would be expected to increase their subsequent memorability. The distinction between Type I and Type II rehearsal is supported by the fact that negative recency effects can be manipulated directly by the task manipulations introduced by the experimenter.

The second related source of empirical support for the distinction derives from studies demonstrating a storage/coding tradeoff. According to virtually all authors including Craik, the information processing system has some maximum capacity to deal with incoming information. Our earlier discussion about capacity becomes more complex with Craik's suggestion that both types of rehearsal monopolize some proportion of this capacity. There is no attempt by

Craik to predict how much capacity each type of rehearsal might require. Indeed, specifying parameters for the capacity needed to process (Type I or Type II) at any given level within the system would be mere guesswork. However, it might be predicted that analyzing an item for elaborate semantic information (Type II rehearsal) would require more capacity than simply reanalyzing a stimulus for low level phonemic information (Type I rehearsal). There is evidence consistent with this general idea. It has been shown that the average time to react to a stimulus is reduced if the item had been presented immediately prior (Hyman, 1953; Keele, 1969; Posner & Boies, 1971). This would seem to suggest that there is a certain amount of facilitation in analyzing a stimulus if it has just been processed in the system. Consequently, Type I processing might be expected, in general, to require somewhat less capacity than Type II processing since the former simply involves re-analyzing for information already extracted.

What empirical implications does this have? Basically, the more the subject engages in elaborative rehearsal of list items, the fewer will be the items for which he can maintain activation. If the subject chooses to simply store items using Type I rehearsal, he will be able to maximize his output from primary memory in the short term. However, if he engages in Type II rehearsal, he will be trading off some of this storage capacity and thereby reduce output from primary memory although, of course, long term memorability will be enhanced.

Evidence for this storage/coding tradeoff may be found in a number of sources. Bellezza and Walker (1974) investigated this issue directly. They presented subjects seven 10-word lists and asked for immediate free recall. A gaming situation was employed in which subjects were told that they would accumulate points for each item recalled. In addition, one group (coding condition) was advised of the FFR test in ~~which~~ additional points could be earned. The other group (storage condition) was not told of this later test. Since recalling items in FFR was much more valuable (10 points) than in IFR (1 point), it was suggested that subjects in the coding condition would be processing items to maximize long-term retention and this might serve to reduce the amount of processing capacity available for short-term storage of items. On the other hand, subjects in the storage condition would have no reason to develop rich elaborate memory traces for long-term retention but rather would utilize their entire processing capacity for the short-term storage of items. An interaction is predicted: subjects in the storage condition should show superior recall on the IFR test but poorer recall on the FFR test. Indeed, these subjects recalled 30% more in IFR than did the subjects in the coding condition but, on FFR, their recall was only half that of the coding group.

In a similar vein, Watkins and Watkins (1974) presented subjects seven lists of words, each randomly varying in length from 8 to 20 items. In one condition, subjects always knew how many items would be in the list (informed group) while in the other condition, subjects only knew list length varied but not how many items would be shown in the present list (uninformed group). The informed group would be able, then, to maximize their performance by simply maintaining the last few items in STM, or primary memory, and then recalling them prior to attempting recall of earlier items. This strategy would not be beneficial for the uninformed group since they never knew how many items were yet to be presented. Therefore, to make a comparison with the earlier study by Bellezza and Walker, the informed group would correspond to the storage condition (with reference to the recency items) while the uninformed group would correspond to the coding condition.

The results showed that the informed group recalled significantly more items in IFR than the uninformed group but that their performance was significantly poorer in FFR. Again, this evidence suggests the storage/coding tradeoffs of Type I versus Type II rehearsal of recency items. Maskarinec and Brown (1974) used a similar procedure in which subjects were presented an unexpectedly short list of only 12 items. Consistent with the Watkins and Watkins findings, the recency effect in IFR was reduced while negative recency in FFR was not only eliminated but in fact was significantly reversed (i.e., positive recency was obtained).

Comparable effects can be found in other situations.

Smith, Barresi, and Gross (1971) found a storage decrement for the last two items of a paired-associate list for a group instructed to generate images of the pairs in contrast to a group instructed to merely repeat the pairs overtly. However, the former group performed significantly better on items in all other list positions. Mazuryk (1974) required different groups to generate semantic associates for each of the last four list items or to overtly or silently rehearse these items during the 3 sec study interval. Those subjects devoting their processing capacity to generating associates showed substantial recency effects in IFR although it was significantly less than that found for the other two groups. Again, though, positive recency was observed in FFR for the associate group whereas the silent and overt rehearsal groups both exhibited significant negative recency. Mazuryk and Lockhart (1974) have replicated this effect under very similar conditions.

In each of these studies, a storage/coding tradeoff was evidenced. Whenever subjects devote more processing capacity to Type II rather than Type I rehearsal, their immediate recall performance suffered, although later recall was concomitantly enhanced. This confirms Craik's prediction based on the differential effectiveness of the two types of rehearsal and thereby provides further support for the distinction.

Comments

It appears, then, that there exists a substantial amount of supportive evidence for the various features of the depth of processing model as it was proposed in 1972. Before discussing the recent modifications to the theory, a number of comments are in order with regard to the earlier versions.

One well documented problem is that there is no a priori definition of depth. That is, Craik provides no basis for determining that one type of information must logically be extracted from a deeper level within the system than some other type of information. Rather, he simply assumes that semantic information originates from a deeper level than does phonemic and that phonemic information is deeper than structural. It does not seem unreasonable that phonemic is deeper than structural if only because it seems that structural information must be extracted from a stimulus array before phonemic information can be made available.

However, the relationship between phonemic and semantic information is less clear. For example, whether "READ" is pronounced with a long or short vowel sound is determined solely by the temporal features which are implicated by the semantic context and preceeding words. The enunciation of other words is clearly determined by semantic context. Thus, "TEAR" can mean a rip if pronounced one way but the liquid output of glands about the eye if pronounced differently. It might also be noted that elaborate structural and phonemic analyses of some words does not help in specifying their meaning. A "BANK" may be a financial institution or the side a river.

A further definitional problem arises when considering some effects of imagery instructions on retention. Morris and Stevens (1974) and Rowe and Paivio (1971) both found that instructing subjects to develop compound images of two or three list items resulted in significantly better recall compared to subjects told only to make an image of individual items. It is not entirely obvious that the former instruction should result in greater depth of processing compared to the latter, particularly in terms of the magnitude of the group differences. This problem of defining "depth" is one which most researchers have simply chosen to live with. Many have accepted Craik's ordering of the levels of analysis at least in part on the basis of the retention characteristics associated with each. To quote Craik and Lockhart, "the comparison of retention across different orienting tasks...provides a relatively pure measure of the memorial consequences of different processing activities" (1972, p. 677). Unfortunately, this renders the definition of depth theory-dependent and thus tautological.

Of course, this is the type of problem from which many theorists have suffered. We noted earlier that the multi-store theorists have yet to provide the theoretical groundwork for deciding how to identify items retrieved from

STM and LTM respectively. More directly relevant to our discussion of depth of processing, though, is a quotation cited earlier from Postman (1964). He suggested that "one may conceive of a continuum of orienting tasks, ranging from those requiring responses maximally favorable to learning to those requiring responses maximally antagonistic to learning." Unfortunately, Postman does not specify in any detail the relative effectiveness of various tasks, with one or two exceptions. Rather, the ordering of each task is again determined by the empirical findings.

A further comment deals with the effect of depth of processing on retention. Rather than suggesting that depth per se is the critical factor, we might speculate that it is the similarity among stored traces which is the critical factor (Bird, 1976; Klein & Saltz, 1976; Posner & Konick, 1966). Thus, early visual analyses of orthographic features of successive items are likely to be much more similar to each other on the whole than later analyses, such as phonemic and semantic. This in itself could be a sufficient basis for the different retention characteristics associated with each type of memorial code. Obviously this is not a novel suggestion (e.g., Shulman, 1971). Posner and Konick's (1966) "acid-bath" model for storage is an example of how such a process might operate. According to these authors, memorial traces degrade each other as a direct function of the amount of similarity existing between them and the length of time they have been in storage. Alternatively, we might suggest that the effect of similarity, and therefore depth, is to reduce the efficiency of retrieval cues in some way.

While Craik does not make reference to such an interpretation in the early statements of depth of processing, he does allude to it in later papers (Craik & Jacoby, 1975; Moscovitch & Craik, 1976). Therefore, we will withold further comments on this matter until the recent revisions of the theory have been considered.

The next two comments on depth of processing are more directly related to empirical observations. It is stated in the theory that subjects who are instructed to learn a list of words will presumably analyze the items to some elaborate semantic level. In this way, retention will be maximized. However, it has often been found that intentional learning subjects who also perform a relatively low level orienting task recall the items much more poorly than subjects who are simply instructed to learn and do not perform the orienting task (Johnston & Jenkins, 1971; Mondani et al., 1973; Till, Diehl, & Jenkins, 1975; Treisman & Tuxworth, 1974). In some cases, this might be understandable. For example, Johnston and Jenkins required one group of subjects to generate and write down a rhyme for each list item presented. Each item appeared for 5 sec during which time the task was to be completed. In the aforementioned unpublished study on directed forgetting by Horton and Petruk, essentially the same procedure was used. In that study, subjects were also restricted to a 5 sec interval to generate an appropriate word. It was very clear to those experimenters that subjects were pressed for time to perform the task. It may be, then, that Johnston and Jenkins' subjects simply did not have time to elaboratively encode items semantically, even though they were instructed to learn the list.

However, other work in which subjects were not under severe temporal constraints also showed a decrement when intentional groups perform lower level orienting tasks (e.g., Mondani et al., 1973). After their subjects had completed a structural task, they were given a fixed interval of 5 sec prior to the onset of the next item.

It seems doubtful that the depth of processing model, as it was formulated in 1972, can account for these findings. The only interpretation which could be offered is that these subjects did not in fact use the opportunity to elaboratively encode the items beyond that required by the orienting task. Of course, this returns us to the tautological definition of depth itself. It will be shown later that there may be provision for these findings in subsequent statements of the theory using the concept of reconstruction.

The second empirical observation for which no provision is made is the depressed retention for items encoded incongruously in the depth of processing paradigm. That is, for those items to which the subject responds "NO" when asked a particular question, severe retention decrements occur quite reliably when compared to items encoded congruously at the same level (e.g., Craik, 1973; Craik & Tulving, 1975). These decrements typically correlate directly with the level of the encoding question but this seems to be primarily a result of floor effects in the recall of items encoded structurally. Schulman (1975) has also called attention to this phenomenon.

Specifically, this disparity in the recall of items encoded congruously and incongruously poses a problem since both types of items must be encoded to the same depth in order for subjects to make a correct decision. Thus, deciding whether "TIGER" belongs to the category "ANIMALS" or "PROFESSIONS" necessitates processing at comparable levels within the system. If the model is correct in saying that the depth to which an item is encoded is the critical factor determining memorability, then there would appear to be no logical reason why the nature of the relationship between the target and encoding question should affect retention.

Again, though, this phenomenon may be explicable in terms of later theoretical developments and we will therefore refer back to it after we have outlined these developments.

Finally, at a very general theoretical level, there are two interesting aspects of the theory which deserve comment. First, it may be noted that there is a very strong emphasis on the encoding and storage aspects of memory. Indeed, in the present review, we have elaborated on the two significant features of the model, namely the correlaton between depth and memorability and the distinction between Type I and Type II rehearsal. Both of these features deal most specifically with encoding and/or storage characteristics.

While this emphasis is clearly evident it has not been to the total exclusion of the role of retrieval. Despite Craik and Lockhart's concluding remark that "no attempt has been made to specify...how items are...retrieved from the system" (p. 682), they do make one or two references to retrieval functions. For example, at one point they suggest that "the form of processing which will prove optimal depends on the retrieval or trace utilization requirements of the subsequent memory test" (p. 678). By this, they are simply saying that different test procedures (e.g., recognition, free recall, cued recall) may demand the use of different types of item information. In an attempt to account for the results of a number of experiments (Dornbush & Winnick, 1967; Eagle & Leiter, 1964; Estes & DaPolito, 1967), they suggest that elaborate semantic encoding may facilitate free recall but may inhibit the encoding of relevant discriminative information for a recognition test (cf. Underwood, 1969). Further, certain types of elaborative encoding may facilitate retrieval in paired-associate (Wicker & Bernstein, 1969) or free recall (Morris & Stevens, 1974) situations while others may not. A number of authors have discussed this matter in a recent book (Brown, 1976) although not all authors have agreed on the hypothesis of qualitatively different types of information being required for different test procedures (see, in particular, Tulving's chapter).

Craik and Lockhart make a second brief reference to retrieval functions. They suggest that "the effectiveness of a retrieval cue depends on its compatibility with the item's initial encoding or, more generally, the extent to which the retrieval situation reinstates the learning context" (p. 678). This statement appears quite consistent with recent remarks by Tulving on "encoding specificity" (e.g., Tulving & Thomson, 1973) and also Martin's description of "encoding variability" (Martin, 1968; 1972). Both of these authors are primarily concerned with the degree to which the encoding conditions are restored at retrieval, making their views very compatible with those of Craik and Lockhart. However, the latter authors do not elaborate beyond this point on the nature of the retrieval process.

Thus, we must conclude that retrieval is left somewhat ill-defined in the early versions of the theory. In the later papers, however, Craik does advance a more complete description of retrieval processes and this will serve as a partial basis for the research to be presented.

As a final point, we would like to emphasize the idea that depth of processing can be viewed as a model of memory based on attributes. The essential idea of the attribute, or feature-analysis, approach is that input may be broken down during encoding for its individual components and subsequent storage and retrieval utilizes these components (cf. Herriot, 1974; Neisser, 1967). Attribute models have been referred to previously in this paper. It can be clearly seen from depth of processing that each level in the processing system leads to the extraction of certain features of the input. The products of each of these analyses are then stored. The critical characteristics of attributes in depth of processing is that they require differing amounts of the processor's capacity to be extracted and maintained and, further, they are lost differentially over time.

Recognizing the importance of attributes within depth of processing simplifies a comparison with other theoretical views, including the multi-store model. Indeed, the latter also derives from an attribute approach to the extent that certain features of the input are analyzed and retained in different memory stores. The attribute approach to memory is unquestionably the overriding paradigm in modern memory theory (cf. Herriot, 1974; Murdock, 1974) and depth of processing is but one manifestation of it.

Revisions to the Theory

At this point, we wish to discuss the various modifications made to the theory as presented in a series of papers (Craik & Jacoby, 1975; Craik & Tulving, 1975; Lockhart, Craik, & Jacoby, 1976). In these recent versions, there have been a number of changes to the original theory in addition to some extensions into areas previously ignored.

In the early statements of depth of processing, it was postulated that information was processed through a sequence of analyses. These processing stages were seen as existing on a continuum, beginning with basic structural analyses and progressing through deeper semantic analyses. In the later versions, most specifically Lockhart et al. (1976), this idea was altered such that now each level in the processing system is represented as a qualitatively distinct "domain" within a hierarchy of analytical stages. The reasons for revising the theory in this manner are strictly logical rather than empirical. That is, as the authors cogently note, "there is no sense in which the physical analysis of...words...shades off into the meaning which the words convey" (Lockhart et al., 1976, p. 78). The rationale appears to be incontrovertible.

This shift away from the idea of a continuum of analyses does not alter the basic concept of depth as it was originally proposed: semantic analyses, for example, are still viewed as "deeper" than structural or phonemic analyses. However, a second dimension of depth is added. Specifically, depth refers not only to the relationship between levels, or domains, but also to the relationships within a domain. Thus, information may be coded to varying depths within a given domain. Examples of this are numerous: whereas we normally perform only a basic structural analysis on written text, a handwriting expert might process the same information for very deep and elaborate structural components. It is this particular aspect of the concept of depth (i.e., depth within a domain for which the authors suggest the term "spread of processing" might be more appropriate. The implication of the term "spread" is that there exists a certain flexibility in the way in which information may be elaboratively encoded within a given domain. In this sense, their use of the term "spread" reflects their attempts to get away from the idea of a continuum, which embodies the idea of a programmed and rigid sequence of analyses being carried out in the system. In the end, however, the term "depth" is retained to refer to processing both within and between domains because it reflects the critical feature of the model: depth in either sense is directly correlated with memorability. Thus, all else being equal, elaboratively encoding information to deeper levels within or between domains will result in increased retention over time.

Regardless of whether the processing system is viewed as a continuum of analyses or discrete domains, it is still maintained in both the earlier and most recent statements of the theory that information extracted at each level in the system is used at subsequent levels. Thus, the products of all the structural analyses are used to extract the relevant phonemic information from the next domain. However, in describing this process, Lockhart et al. extend their ideas to include the idea of expectancies. This is illustrated in the following passage. Note that their use of the term "structural" here is not meant to refer to the particular level of analysis within the system.

> A structural description of the input pattern is formed within each analytic domain. At each level of analysis, the structural descriptions formed within one domain serve as the input to the next domain via a set of mapping rules. A slightly different way of phrasing these ideas is to say that each level of analysis provides evidence which is used to confirm (or reject) the structural description of the hypothesized patterns at the next level. This second way of describing the process may be preferable in that it stresses the notion that structural descriptions at any level are as much a product of expectancies and past learning as they are products of the current stimulus input. Further, descriptions of very probable events will require only minimal confirming evidence from preceding levels, since their structural descriptions have been largely performed in anticipation. (Lockhart, Craik, & Jacoby, 1976, p. 78; emphasis is original)

We might suggest that the latter ideas are rather more than just "slightly different". What is actually being offered here is a type of hypothesis-testing model in which information is generated within the system prior to the actual processing of the input. This view is not dissimilar to that of other authors, as is pointed out. For example, Treisman (1964) has suggested that, on the basis of prior inputs, certain events may be "expected" more so than others. This results in the thresholds of the "dictionary units" for these expected events being lowered relative to all other possible events. This hypothetical process ensures that high probability material will be perceived and that other irrelevant information will not interfere substantially. Norman's (1968) "pertinence" selector operates in essentially the same manner.

Whether or not this sort of hypothesis-testing approach will be viable is perhaps debateable. Fortunately, however, we need not involve ourselves in this matter, even though it is a recurring theme in the Lockhart et al. paper. In the typical human memory experiment, and in particular those to be presented later, expectancies serve no meaningful role since subjects are presented lists of random words. When subjects are shadowing meaningful prose passages, as they do in the research referred to by Treisman and Norman, expectancies may be more important.

The effects of practice, or familiarity, are one aspect of processing which were not dealt with previously but are important considerations in the modified versions of the theory. This issue is discussed at length by Lockhart et al. (1976) and also to some extent by Craik & Jacoby (1975). Basically, it is postulated that the actual number of analytical operations carried out on an individual item decreases as a function of practice or experience with the item. Indeed, the authors point out that it may be possible to bypass or eliminate an entire processing stage with substantial amounts of practice. They use the example of highly skilled readers who may be able to process text without invoking any phonemic analyses, despite their early training which heavily emphasized phonemic coding.

A direct result of the fact that the amount of processing carried out at various levels decrease as a function of practice is that there is a concomitant decrease in the amount of information available for representation in the memory trace. Thus, the gain in efficiency of coding operations which occurs with practice is translated directly into a loss in memorability since less information is extracted and stored. The effect of practice in terms of subsequent memorability, then, is essentially the same as it is for high expectancy events, as may be surmised from the earlier quotation: a relatively meagre trace is established in both cases and this leads to a decrease in retention. It may be noted that the idea of continuity between perceptual and memorial processes is strictly maintained here, since what is perceived determines what is stored in the memory trace.

The representation of practice in the system has a second important effect. Information which is highly practised or very compatible with deep level processing (these terms seem to be used synonymously) may be automatically analyzed rather extensively: "it may be impossible to prevent processing occurring at a deeper level" (Lockhart et al. 1976, p. 79) than that necessitated by the task. In other words, they are suggesting that analysis of unfamiliar material may require conscious processing but analysis of material with which we have had much experience, such as common words, may proceed automatically to a relatively deep semantic level. This may occur despite the nature of the current task, a good example being the Stroop effect.

The authors are careful to point out that their use of the term automatic processing implies not only that there is a certain inevitability about the processing which will occur but also that conscious attention need not be invoked. Thus, they specifically say that "such analyses require little conscious attention to be carried out effectively" (Craik & Jacoby, 1975, p. 174).

So far we have summarized the view that practice results in the deletion of certain coding operations and that, with extended practice, coding (of verbal material) to a semantic level may occur automatically and without cognitive intent on the part of the subject. The memorial consequence of such highly skilled, automatic processing is what Craik and Tulving refer to as the "minimal core encoding" (1975, p. 290). The term seems to derive from the idea that a very limited number of analyses are likely to be enacted at each level. Therefore what enters into the memory trace is a minimal representation from each level. However, the nature of the task may require a substantial analysis at, say, a structural level. Therefore, the subject may need to consciously return to a low level of processing so as to carry out the elaborate structural analyses which did not occur during automatic processing of the item. In this sense, then, the core encoding derived initially may be elaborated at any given level of processing. The implications of a minimal core representation for retention will be discussed and tested in two experiments presented later.

The first significant change to the theory, then, involves the hierarchical arrangement of processing domains, this in contrast to the earlier idea of a processing continuum. At a more general theoretical level, Craik and his colleagues have also chosen to embrace the distinction between episodic and semantic memory (Tulving, 1972). Since Craik suggests that his use of the terms is somewhat different from the original description offerred by Tulving, it seems appropriate to describe Tulving's view in brief.

Essentially, semantic memory is the system of "cognitive structures" in which knowledge is represented. There are a number of different aspects to these cognitive structures, but, for present purposes, the most significant are the rules and procedures used for analyzing and interpreting incoming information and also the organization of language. Tulving attributes to semantic memory all information or knowledge which is not stored in the context of its spatio-temporal reference. Episodic memory, on the other hand, is the storage system for all the information retained about a person's individual experiences. A condition of storage in episodic memory is that spatio-temporal attributes are available. Tulving uses the term autobiographical events to refer to episodic traces.

For our purposes, it is of interest to note that "it is...possible for the episodic system to operate relatively independently of the semantic system" (Tulving, 1972, p. 386). However, it is further acknowledged that "semantic memory is the memory necessary for the use of language" (p. 386). Thus, in the typical human memory experiment "where to-be-remembered units are meaningful words that refer to concepts stored in semantic memory, the information in semantic memory may be used at the time of the input of the information into the episodic memory store" (p. 390). Clearly this must be so since the rules and procedures for encoding structural, phonemic, and semantic information must be stored in some long-term system (cf. Smith et al., 1971). An episodic trace, then, cannot be established in such an experiment without input having been first processed through the semantic system.

Craik suggests that his use of the episodic-semantic distinction varies from that of Tulving. Rather than allowing that episodic memory may operate "relatively independently" of semantic memory, Craik chooses to emphasize the "very close interdependence" (Craik & Jacoby, 1975, p. 174) of the two systems. Indeed, he suggests that "the semantic-episodic distinction actually refers to aspects of one system rather than two distinct systems" (Craik & Jacoby, 1975, p. 174).[6] Based on the previous very brief, selective review of Tulving's views on the episodic-semantic distinction, it appears that there may be no real points of disagreement. This is especially evident in our present concern with verbal memory experiments. Tulving's use of the term "independence" in reference to the functioning of the two systems may have been unfortunate to some degree since it seems to refer to aspects of the systems which are not relevant to memory research in general.

---

[6] We must be careful to take cognizance of Tulving's own views on the idea of distinct systems. In his own words, "I

We are forced to conclude, then, that Craik and his colleagues have not really cast the episodic-semantic distinction in a different light than has Tulving. But a more important theoretical problem must be resolved in Craik's use of the terms.

Craik and Jacoby emphasize that, while they see episodic and semantic memory as referring to two dimensions of the same system, they feel that "the heuristic and conceptual advantages of regarding them as separate, outweigh arguments for a unitary system" (p. 174). Unfortunately, they do not specify why this is the case. Indeed, we may ask exactly what theoretical or heuristic gain is in fact made by positing an episodic-semantic distinction. It is contended here that there are no

---

will refer to both kinds of memory as two stores, or two systems, but I do this primarily for the convenience of communication, rather than as an expression of any profound belief about structural or functional separation of the two" (Tulving, 1972, p. 384). On this issue, then, there may be no disagreement.

"heuristic and conceptual advantages" to it. This certainly seems true in terms of the encoding of information in all the research we have been, and will be, considering. Similarly there appears to be no theoretical utility in terms of retrieval. In summary, we remain uncertain as to why episodic and semantic memory have been differentiated within the context of depth of processing.[7] If there is good logic behind it, the authors have not made it evident.

Perhaps the most interesting extension of the depth of processing model is the recent concern with retrieval. In the previous discussion, we noted that very little was said about retrieval processes in Craik and Lockhart's (1972) early statement of the theory. They currently envisage three major modes of retrieval, each of which we will document in turn. As might be expected, the characteristics of each differ and, in addition, each plays a role in various "short-term memory" phenomena, such as recency effects (e.g., Glanzer & Cunitz, 1966; Bjork & Whitten, 1974).

---

[7] We are not convinced that the distinction is really necessary in any context. If Tulving wishes to show simply that the coding operations carried out on input are theoretically distinguishable from the products of these operations (i.e., the conscious percept and, possibly, the memory trace) and that some stored information retains its autobiographical reference, he need not have parsed memory into two distinct systems. The first issue is undoubtedly pretheoretical for most authors (but see Kolers, 1973, and Neisser, 1967). On the matter of the second issue, Tulving's views may be oversimplified. We might suggest that it would be much more realistic to regard autobiographical information as being of a continuous nature, rather than all-or-none.

The first type of retrieval involves simply the output of information which currently exists in consciousness. Thus, the products of the various levels of analyses which have entered consciousness may be "retrieved" from there directly at any point in time as long as the interval between presentation and test was used to continually attend to the information. Essentially, they are inferring Type I maintenance rehearsal as the principal basis for this type of retrieval although, of course, Type II rehearsal also involves continued attention to the information, as we have previously noted. There is an obvious affinity between retrieval from consciousness and retrieval from the theoretical "rehearsal buffer" (Atkinson & Shiffrin, 1968) since in both cases the information is currently "active" in the system. Similarly, there is an empirical correlate to this in the overt rehearsal literature (e.g., Rundus & Atkinson, 1970; Brodie, 1975).

Mandler (1975) has remarked on a number of interesting issues with respect to the idea of consciousness and conscious attention. In particular, he has offered the conceptual argument that 'the information that, so to speak, can be "read off" the contents of consciousness is not memorial as such' (p. 237). His reason for saying this is based on the idea that "retrieval usually implies retrieval into the conscious field" (p. 237) and therefore material currently in consciousness need not be retrieved. He is in essence co-defining memory as retrieval. This argument, though, is purely conceptual in form. Indeed, we may choose to define memory alternatively as the output of information which is not currently available from the environment or as the output of information directly from the cognitive system without reference to the environment (whether or not the information is available there). In this way, the output of conscious material would classify as a memory or retrieval phenomenon. Lockhart et al. (1976) have similarly attempted to show that the problem surrounding the issue of retrieval is definitional and on that premise have chosen to avoid the term. We shall have a few more comments on the matter of consciousness in general at a later point.

While Craik has relatively little to say about retrieval of information residing in consciousness at the time of retrieval, it seems to be implied that no forgetting of this information occurs. Obviously, then, there are also no forgetting effects attributable to the nature of the code (e.g., phonemic, semantic).

The second retrieval mode postulated by Craik and his colleagues is called the backward search or scanning process. There are three critical characteristics of this retrieval mode: it can be used only for recent entries into episodic memory; the retrieval cue is used to discriminate the target from other items in memory; there are no forgetting effects due to the type of information stored.

In regard to the items which may be retrieved by backward scanning, Craik and Jacoby have specified that "the search process proceeds backward from the present and becomes rapidly less efficient as increasingly more items intervene between presentation and test" (1975, p. 176). In other words, the last entries into episodic memory are scanned in reverse order for the target item. Exactly how far back into episodic memory this process can operate is not specified by Craik. On the basis of the data to be reported later, however, it would seem to vary as a function of the type of retrieval information provided at the time of test. Thus, simply requiring subjects to recall the last of a series of lists normally provides little basis for discrimination; specifying that the target item is related to a retrieval cue on some attribute is somewhat more effective; asking the subject to make a "yes/no" recognition judgment provides, generally, the most effective information. Thus, as the overall amount of information given at the time of testing increases, the number of intervening items over which backward scanning can be effective also increases. The evidence cited by Craik in support of this, and comments on it, will be discussed later. The essential aspect of backward scanning, though, is that it is used to discriminate items in memory, rather than actually retrieve them: "the retrieval information is not used to provide access to the trace in any sense, but is used merely to select the target items from other items" (Craik & Jacoby, 1975, p. 176). According to this, then, the target items must be highly accessible in a sense not dissimilar to that postulated by most multi-store theorists in their description of items in STM. The term "retrieval information" should not be confused here even though retrieval does not actually occur when backward scanning is employed. It simply refers to the information provided by the experimenter specifying to the subject what the target information is.

The final characteristic of backward scanning is its insensitivity to the nature of the retrieval information used to select the target items. Specifically, Craik is suggesting that, all other factors (e.g., similarity of the episodic traces) being equal, semantic, phonemic, or any other retrieval information will be equally effective for recall.

It was this latter characteristic which led Jacoby (1974) to originally discriminate backward scanning from the final retrieval mode, that being reconstruction. He observed that under some conditions differential forgetting did not occur as a function of the type of information provided at the test. However, as the delay between presentation and test increased, these differential effects began to appear. Long-term retention differences as a function of the nature of the retrieval information are indeed the rule, but exceptions like that of Jacoby's have been noted by other researchers (e.g., Bregman, 1968; Shulman, 1970). These data will be discussed more fully when we deal with the evidence for the various features of the updated depth of processing model.

When information in episodic memory is too dated to be accessed by backward scanning or when the retrieval cue is insufficient to permit discrimination among the targets, subjects are required to use the final retrieval mode, reconstruction. The basic feature of reconstruction is that the subject attempts to reinstate "the original perceptual encoding of the event" (Craik & Jacoby, 1975, p. 176) upon presentation of the retrieval cue or information. There is an obvious similarity between how the reconstruction process operates and the encoding specificity principle (e.g., Light & Carter-Sobell, 1970; Tulving & Thomson, 1973).

Craik and Jacoby suggest that there are two important factors involved in reconstruction: "the reconstructive activities involve the cognitive structures[a], as did the initial encoding, and are guided and constrained both by the structure of semantic memory and by feedback from the episodic trace itself" (1975, p. 176). By the first part of this statement, the authors are simply saying that the retrieval information is processed just as any other information. Obviously, then encoding of the retrieval information must conform to the structure of semantic memory in the sense that all the analytical routines are stored there. In suggesting that reconstruction is also controlled in part by "feedback from the episodic trace", they seem to be implying that psychological effects such as tip-of-the-tongue (Brown & McNeill, 1966) or feeling-of-knowing (Blake,

---

[a] By "cognitive structures", they are referring to the analytical routines in semantic memory (see Craik & Jacoby, 1975, p. 174).

1973) phenomena become operative. Accordingly, they talk about "feelings of partial recognition as the developing percept approximates the structure of the episodic trace" (Craik & Jacoby, 1975, p. 176). The result of these feelings of partial recognition is that "'positive feedback' will result and further reconstruction along the same lines is encouraged" (Lockhart et al., 1976, p. 85). Thus, with the reconstructive processes progressively "homing in" (Jacoby, 1974, p. 494) as more features of the episodic trace are developed, the feelings of familiarity become increasingly strong. Lockhart et al. summarize their view by saying that "this 'guided reconstruction' is seen as a servomechanism in which feedback from the target controls the reconstructive processes" (1976, p. 85).

The authors have provided a more explicit statement as to how reconstruction operates in the depth of processing paradigm. They suggest that, "at retrieval, it is likely that the subject uses what he can remember of the encoding question to help him retrieve the target word" (Craik & Tulving, 1975, p. 284). In other words, recall of the specific encoding question seems to be critical in the role of reconstruction: it appears that this aspect of the encoding context must be retrieved. This operational definition of reconstruction as it relates to the depth of processing paradigm will be important in the research to be reported. It might be noted in addition that this interpretation can be assumed to apply to standard memory paradigms in which encoding questions are not asked.

However, since it is impossible to determine exactly what is coded in these unstructured situations, little can be said about what it is in particular that the subject is using (if and) when he reconstructs the original encoding.

To facilitate further understanding of reconstruction, Craik compares it to the generation-recognition approach (e.g., Bahrick, 1970) and search or scanning models of memory (e.g., Sternberg, 1966). Whereas generation-recognition models imply that an entire target is retrieved by the system, the reconstruction process operates on partial information. Thus, the feelings of familiarity develop as an increasing number of target fearures are accessed. Search models suggest that "stored events are examined...until the desired information is located" (Lockhart et al., 1976, p. 85). In reconstruction, feedback from the memory trace helps the subject to retrieve the trace. In this sense, the trace need not actually be accessible to the subject directly (i.e., he need not consciously know, or be able to recall, the item) before reconstruction can commence whereas, in search models, the implication is that the targets are in conscious awareness and need only be compared with the probe information.

Lockhart et al. also provide an interesting view of the processing differences in recall and recognition testing procedures. Rather than suggesting that recall and recognition involve qualitatively different processes (e.g., Kintsch, 1970), they hypothesize that these two types of tests differ only in the nature of the retrieval information provided and what is to be retrieved from memory: "in recognition, the stimulus is re-presented and the system has to reconstruct the context; in recall, some aspects of the context are re-presented or referred to and the system has to reconstruct the stimulus" (Lockhart et al., 1976, p. 85). There is a substantial similarity between this view and that of other authors who suggest that recall and recognition are simply components of a continuum. For example, Tulving (1976) has postulated that "recognition and recall differ only with respect to the exact nature of the retrieval information available to the rememberer" (p. 37).

Reconstruction differs significantly from the other two modes of retrieval with respect to the effects of how various types of information affect retrieval. When information is output directly from consciousness or via the backward scanning process, semantic, phonemic, or any other type of cues are equally effective. However, "deeper, semantic information is usually much more effective in the process of reconstruction" (Craik & Jacoby, 1975, p. 176) than is lower level retrieval information. It is with this postulate that depth of processing accounts for the many findings of differential forgetting as a function of the qualitative nature of the encoding. The empirical support will be summarized below.

According to Craik, each of these three retrieval modes may be a factor in the interpretation of typical short-term memory phenomena. The prime example which they discuss is the recency effect. In a standard free recall paradigm, the last item presented is almost invariably recalled correctly given the recall is tested immediately after input or if rehearsal (without interference) is permitted during a delay interval. Even if list length is varied and the subject is unaware of when the last item is to appear, recall of this item may still correspond to that of informed subjects, with performance at a level greater than 90% (e.g., Watkins & Watkins, 1974). In these and other situations, it may be assumed that the last item is probably still in consciousness and therefore there is no problem in the output of it. Once an item is dropped from conscious awareness (i.e., is no longer "attended"), it must be retrieved by some other means, although Craik and Jacoby leave open the possibility that some feature or attribute of the item may remain consciously available (cf. Rumelhart et al., 1972; Weist, 1972). Various distractor tasks (cf. Atkinson & Shiffrin, 1971; Peterson & Peterson, 1959) may serve to eliminate the last list item from consciousness and thereby yield at least some short-term forgetting.

The backward scanning process also plays a significant role in recency effects. Specifically, the observation that probability of recall decreases rapidly across the last six to eight input positions (cf. Murdock, 1962) is attributed to the declining efficiency of backward scanning. Presumably, since the recall function levels off for previous input postitions, reconstruction must be initiated.

For these items, backward scanning is apparently too inefficient to be continued. It is a common observation that recency functions do not differ over a variety of conditions (when corrected for asymptote level), including the nature of the encodings (Glanzer, 1972). This finding is expected on the premise that backward scanning is unaffected by the nature of the encoding and since this process, according to Craik, involves only the discrimination of targets and not actual retrieval.

Finally, rather typical recency effects have been observed (e.g., Bjork & Whitten, 1974; Tzeng, 1973) under experimental conditions which would seem to preclude short-term memory effects according to traditional multi-store theories. In these studies, subjects were presented a list of words or pairs of words. After each item appeared, the subject was engaged in a distractor task for a minimum of 10 - 12 seconds. After all items had been presented in this manner, subjects were asked to recall the list following a minimum delay of 30 seconds of the distractor task subsequent to the occurrence of the last item. Clearly, these delay intervals should have eliminated or at least reduced the amount of information available in STM at the time of testing (e.g., Waugh & Norman, 1965; Atkinson & Shiffrin, 1968). Despite this, recency effects still obtained.

Craik and Jacoby suggest that these findings may be described in terms of long-term recency effects attributable to "the declining effectiveness of the reconstructive processes as the event becomes remote" (1975, p. 180). Unfortunately, it is not apparent why performance asymptotes for middle input positions under standard free recall conditions. If reconstruction is sensitive to recency factors, then it seems theoretically consistent to predict performance decrements across these asymptotic positions. We will comment on this matter later in our discussion of the empirical separation of backward scanning and reconstruction processes. Craik and Jacoby employ the same logic to account for the common finding that final free recall also exhibits recency effects. That is, the last lists presented are recalled better than earlier lists.

In our earlier discussion, we suggested that the depth effect might be explained in terms of differential interference. Thus, the structural analysis of any two given items is much more likely to overlap by chance than is, say, the semantic analysis and, consequently, the stored traces for each item are more likely to share lower level information than deeper, semantic information. It could be, then, that this differential similarity among the various attributes of memory traces results in some form of "trace interaction", to borrow Runquist's (1975) term. In this sense, it would be argued that items having overlapping attributes actually become increasingly less available (cf. Tulving & Pearlstone, 1966) over time (although the empirical support for this hypothesis is not compelling). We previously cited the example of Posner and Konick's

(1966) acid-bath model in this context.

As an alternative interpretation of the depth effect, it was suggested that perhaps similarity serves to reduce the efficiency of the retrieval cues. This is in fact the interpretation which Craik has recently elaborated in terms of retention differences found with the reconstruction retrieval mode. The following extracts from Lockhart et al. will illustrate this view.

> Craik and Lockhart argued that deeper, semantic processing yielded a more "durable" code. An alternative view, which fits more readily into the present scheme, is that all encoded events are equally durable...but that some traces become impossible to access because they are not distinctive but similar to many other events. (p. 89)

Also,

> When a particular pattern of encoding operations is induced by the test stimulus, all episodic traces of the pattern are contacted....If there are many traces of the pattern, a "familiar" encoding will be achieved easily...but the stimulus will not elicit recognition of a specific previous instance (since many different contexts are competing for conscious awareness). If the pattern was unique or distinctive, however, only one or a few traces of such a pattern exist - now if the retrieval information...is specific enough, the previous traces will be contacted. (p. 83; emphasis is original)

The latter exerpt is in reference to recognition testing in particular, but the idea may be generalized to recall.

These statements seem to coincide with the view that similarity is acting to reduce the efficiency of the retrieval information. Thus, if the subject attempts to reconstruct a previously presented item on the basis of some low level information, he is very likely to encounter substantial interference from other items sharing that feature.

This idea is not dissimilar to Runquist's (1975) "associative interference", although in a somewhat different context. According to Runquist, when a stimulus item is presented on a paired-associate test trial, it "simultaneously activates the correct association and one or more incorrect associations" (1975, p. 144), the number of incorrect associations activated being a direct function of the similarity among list responses. Thus, it is the fact that various responses overlap on one or more features that results in many responses being aroused concurrently.

Craik's view varies from Runquists' in that the latter assumes each of these high similarity responses becomes consciously available to the subject more or less directly whereas Craik assumes that the more items which are similar, the greater the difficulty in gaining conscious awareness to any of them (see, in particular, their reference to a "limited energy resonance model", Lockhart et al., 1976, p. 83). More interestingly, these two authors agree with the idea that the subject requires additional discriminative information before a correct response can be achieved. This discriminative function is typically assigned to the "more distinctive and unique" semantic information in the depth of processing model.

We may now interrelate the roles of reconstruction and differential interference at retrieval as they relate to the depth effect. In the typical experiment using the depth of processing paradigm, subjects are presented a series of questions about the target items. Often the structural questions are all of the sort "Is it in uppercase?" or "Does it contain the letter 'e'?" Clearly, recalling either of these questions will be of little benefit if 20 or more different targets were encoded with such questions. On the other hand, phonemic questions such as "Does it rhyme with _____?" and semantic questions such as "Is it a type of _____?" are almost invariably unique for each word. Thus, recalling each of these types of questions is less likely to cause interference in retrieving or reconstructing the target because of their unique relation to a single target. Unfortunately, this overall interpretation only accounts for why structurally-encoded items are more poorly recalled than phonemically- or semantically-encoded items but does not account for why the latter two also differ significantly. Craik and Tulving propose that "although each target word receives a different rhyme question, phonemic differences may not be so unique or distinctive as semantic differences" (1975, p. 285). This is not particularly satisfactory.

However, the analysis does lead to an interesting prediction about the effects of having multiple targets encoded in the context of the same question. If semantic questions are in some way more unique or distinctive than phonemic questions and both are more unique than structural questions, then it should be possible to manipulate the depth by varying the number of targets encoded with the same encoding question at each level in the system. Thus, if lower level questions simply do not exhibit as unique a relation to the target as deeper, semantic questions, then pairing the same phonemic question with multiple targets should result in less of a decrement in performance than pairing the same semantic question with multiple targets. Moscovitch and Craik (1976, Experiment 2) performed this experiment using cued recall and found that, while the depth effect still obtained, it was substantially reduced. Significantly, there was virtually no decrement for phonemically-encoded items when compared with the standard condition in which all encoding questions are nominally unique. There was, however, a substantial drop in performance for both semantic conditions tested in which multiple targets shared the same encoding question.

In another experiment (Craik & Tulving, 1975, Experiment 8), subjects were presented a series of target words each of which was coded with a structural, phonemic, or semantic question. The number of questions from each level of coding varied between subjects, but ranged from 4 to 40. It was hypothesized that, if uniqueness of encoding is a significant variable in determining the magnitude of the depth effect, then the fewer the number of words encoded at a given level (and therefore the greater the uniqueness of each in memory), the greater should be the retention of these items. Using a modified yes/no recognition procedure in which subjects were required to select a fixed number of responses, it was found that the number of questions asked at any given level had no consistent effect on retention. While no statistics are reported, it is clear that there was absolutely no effect on retention of structurally-encoded items. Items encoded phonemically or semantically gave some hint of a performance increment, at least for the congruent encoding conditions, as the number of questions asked at each level of coding decreased. Ceiling effects militate against a firm conclusion on this matter.

An alternative interpretation of these findings might be based on the idea expressed earlier with regard to differential interference at storage. Since some relatively low level information (e.g., structural) must necessarily be encoded and thereby form part of the memory trace for all items presented, it is possible that maximum storage interference was operative even before additional structural information was extracted for the purpose of responding to the encoding question. Thus, varying the number of structural encoding questions presented during study might conceivably have no significant influence on the amount of intereference in effect during storage.

We would like to suggest at this point that both storage and retrieval interference may be factors in the depth effect. Thus, differential storage interference might account for the hypothesized durability of traces coded at different depths within the system. Retrieval interference can apparently be implicated in some situations, such as that described by Moscovitch and Craik. While it seems entirely reasonable to offer the compromise that both factors are important, it may not be a particularly useful beginning for empirical investigations since localizing an effect in the storage or retrieval phase of memory may be experimentally insurmountable (cf. Murdock, 1974). This is especially true if, in accord with Craik and his colleagues, one views recognition and recall testing procedures as inadequate for making the storage/retrieval distinction.

The final issue to be discussed under the heading of revisions to the depth of processing model deals with the matter of congruity. Specifically, we noted earlier that the original versions of the theory could not account for (and did not reflect on) the fact that target items encoded incongruously (i.e., with an encoding question leading to a "no" response) were much more poorly retained than items encoded congruously to the same depth. If in fact depth was the only critical variable determining memorability, as the model hypothesized, then all items coded to the same depth should be equally well recalled.

Craik and Tulving address this issue. They suggest that, memorability will be a joint function of the depth to which the item is coded and the degree to which "the encoding question and the target word can form a coherent, integrated unit" (Craik & Tulving, 1975, p. 281).[9] It is clearly implied, but never actually stated, that subjects attempt to retrieve the encoding questions during recall and from there engage in reconstruction of the targets. With the example of cued recall, they suggest that "representation of part of the encoded unit [i.e., the encoding question] will lead easily to regeneration of the total unit" (p. 291). In free recall, it may be presumed that subjects supply themselves with the encoding questions if reconstruction is to occur. If in fact this is the process subjects use to retrieve targets, it is obvious that congruent relationships are superior in part due to the restricted subset of possible alternatives available. Thus, asking a subject a structural question such as "Is it in uppercase?" Or "Does it begin with the same letter as DOG?" does not functionally restrict the alternatives. Consequently, neither positive nor negative responses to these types of questions should lead to improved performance.

---

[9] They actually parse the latter factor into a) the degree of "encoding elaboration" attained and b) the potential of the

Other authors have postulated essentially the same reason for this finding. For example, Schulman (1974; 1975) has discussed the idea of a "relational encoding"[10] between

---

encoding question to lead to the target word. Based on their discussion, it is difficult to see exactly how one theoretically or empirically separates the two since degree of encoding elaboration of the target covaries directly with the strength of the relationship between the question and the target. For example, they say that "for positive responses the encoding question can be integrated with the target word and a more elaborate unit formed" (p. 291). Similarly, "positive responses would be integrated with the question and thus...formed more elaborate encodings which supported better retention" (p. 283). (There is an exception to this rule - see text.)

the encoding question and the target. He presents data from an experiment using a minor variation of the depth of processing paradigm in which congruously encoded targets were recalled much better than those encoded incongruously (Schulman, 1974). He summarizes these findings by saying that "the pattern of results shows a large and pervasive memorial advantage of congruity, arguably because a congruous query, unlike an incongruous one, fosters a relational encoding of [target] and [question]" (Schulman, 1975, p. 48). Attributing the performance differences as a function of congruity to the relational encoding clearly implies that both elements must be recalled in the congruous condition. Begg's (1972) idea of a "coherent, integrated unit" is not dissimilar from that of Craik or Schulman, although he was specifically interested in interpreting the memorial superiority of images over verbal strings.

Evidence

Three features of the updated depth of processing model which are considered important here are those dealing with the automatic processing of information, and the two major retrieval modes, backward scanning and reconstruction. These aspects of the theory are significant in the interpretation of the research to be reported later and, therefore, we will present here some of the supportive

---

[10] This is defined simply as that feature, or attribute, which links the question and target, be that relation phonemic, semantic, or whatever.

evidence relating to each.

There is a substantial amount of evidence from a variety of research paradigms which lends support to the idea that highly familiar material (e.g., common words) is automatically processed at a number of different levels of processing. In some of these paradigms, this automatic processing facilitates performance but in others it hinders. One example previously referred to was the classic Stroop effect. Subjects are simply presented a list of colour names (RED, BLUE, etc.) and asked to verbally identify, as quickly as possible, the colour of the ink in which the word is printed. If the colour name and the colour of the ink are different, subjects exhibit substantial interference. This implies that subjects cannot avoid processing the name of the word itself in order to identify a more structural characteristic.

Using a paired-associate task, Nelson and Borden (1973) observed that high phonemic similarity among stimulus members significantly interfered with learning in contrast to little or no such similarity. This was despite the very high semantic relationships between each stimulus-response pair (e.,g., HAM - EGGS). Thus, the phonemic similarity among items in this experiment interfered substantially with what appears to be a very simple semantic association task.

In a number of dichotic listening experiments, it has been shown that semantic information on the unattended channel has important effects on the processing of the attended or shadowed message. For example, Lewis (1970) found that shadowing performance was decremented when semantically similar words occurred simultaneously on the two channels. Lackner and Garrett (1973) found that an otherwise semantically ambiguous message on the attended channel could be resolved by presenting appropriate semantic information on the other channel. Treisman (1960) presented subjects a continuous semantic message on one channel and required them to ignore the other channel. During shadowing, the two messages suddenly switched channels. As might be expected if subjects were processing semantically on the so-called unattended channel thereby maintaining the continuous semantic content of the message, these subjects reported that they were unaware of having made the switch. Finally, Corteen and Wood (1972) observed a galvanic skin response to words in an unattended channel which was comparable in magnitude to that when the same word occurred on the attended channel. In all of these experiments, it seems rather convincing that subjects must have been processing to a relatively deep, semantic level on the unattended channel even though normally there is no apparent effect on performance on the primary task.

It may be noted here that many authors have dealt with the matter of automatic processing of information. While we do not wish to comment on this issue at length, it is of interest to note how other authors have conceptualized it. Some have agreed that a minimal semantic encoding of words will normally occur without intention on the part of the subject (e.g., Norman, 1969a; Smith & Groen, 1974) although they attribute no automatic storage of this information over any period of time. Their views coincide with much of the data cited above which reveals major effects of apparently automatic semantic processing of unattended input although retention is minimal (e.g., Moray, 1959; Norman, 1969b).

Posner (Posner & Warren, 1972; Posner & Snyder, 1975) suggests that a substantial amount of processing may occur without intent. In contrast to some authors, though, he wishes to define automatic processing not simply as that which occurs without intention. He extends his definition of automatic processing to include the ideas that no conscious awareness results, there is no "interference with other ongoing mental activity" (Posner & Snyder, 1975, p. 56), and there is no storage of the products of these analytical operations (Posner & Warren, 1972, p. 34). A number of recent papers by Shiffrin and his colleagues (Schneider & Shiffrin, 1977; Shiffrin & Grantham, 1974; Shiffrin & Schneider, 1977) offer a view of automatic processing very similar in some respects to that of Posner. Specifically, they focus on the ideas that automatic processing does not typically result in conscious awareness and that it does not require any amount of the limited capacity processing of the system. They say little about memory for this type of information but it may easily be implied that they agree with Posner on this matter.

In general, we may summarize by saying that Craik's views of automatic processing probably differ in some ways from that of other authors. While we must be careful to point out that Craik is not perfectly clear on some of these issues, it seems legitimate to characterize his views on automatic processing as 1) requiring little or no conscious attention, 2) resulting in a conscious percept, and 3) resulting in a "minimal core encoding" in the memory system. The third feature seems to be a logical corollary of the hypothesized continuity between perception and memory. It is this feature which will be of particular interest in two of the studies to be reported. For present purposes, though, we might just take note of the difference of opinion on whether a memory trace results from automatic processing and also the data cited above suggesting in fact that no meaningful memory trace has been evidenced in situations reflecting what seem to be conditions of automatic processing.

The other two aspects of the updated depth of processing model for which we will present relevant data are the two retrieval modes, backward scanning and reconstruction. Earlier we referenced a few studies which showed no retention differences due to the nature of the encoded information (Bregman, 1968; Jacoby, 1974; Shulman, 1970). Bregman presented his subjects a type of continuous cued recall situation in which a series of words was presented one at a time. Throughout the list, cues were presented intermittently. The cues of particular interest to us are those providing structural (first two letters), phonemic (a rhyme), or semantic (category membership) information. With lags of 1 to 96 intervening items (both study and test intermixed), Bregman found no significant retention differences due to these three types of cues. A study by Shulman (1970) is often cited as evidence for no differential loss of phonemic and semantic information over short retention intervals. Shulman presented his subjects a list of 10 words followed immediately by one of three types of probes. These probes were either identical to one of the list items, or were homonyms (e.g., cereal-serial) or synonyms (e.g., nation-country) of one item. The subjects' task was to indicate "yes" or "no" as quickly and as accurately as possible whether there was an item presented having one of these specified relationships with the probe word. Subjects did not know what type of relationship was to be tested until the entire list had been presented. Two results are of interest. First, reaction time to respond to identical probes was the shortest and was longest for synonym probes. Correct recognition, however, was equal for identical and homonym probes but much lower for semantic probes despite the latter having the highest false recognition rate. Shulman concluded that, on the basis of these and other data showing no differential loss over time for each type of probe, the nature of the encoded information does not affect forgetting rates. This is in line with Bregman's findings.

This study could be interpreted in another way, though. It could be argued that subjects simply rehearsed or maintained phonemic information during the presentation of the list. Then, when a semantic probe was presented at test, they processed the stored items for relevant semantic information in order to meet the task demands. This explanation of the data is supported by the finding of longer reaction times to synonym probes. Also, since processing capacity would be required during the test interval in order to extract the necessary semantic information, some memory loss might be expected (see earlier comments on the storage/coding tradeoff). This would account for the reduced recognition performance for synonym probes in contrast to identical and homonym probes.

This interpretation of Shulman's data seems at least as plausible as that presented in the original paper. Because of this, we do not consider this experiment as providing critical evidence for Craik's backward scanning retrieval mode.

In a study reported by Craik and Jacoby (1975), subjects were tested in a continuous recognition procedure much like that of Bregman (1968). One significant difference was that in the more recent study, list items were presented in the context of encoding questions which required either structural, phonemic, or semantic information whereas words were presented in isolation in Bregman's study. Further, Bregman presented either structural, phonemic, or semantic cues at test and required subjects to indicate whether a related study item had occurred. Craik and Jacoby simply presented target and distractor items and asked for a yes/no recognition judgment. As in the earlier study, recognition performance did not differ as a function of the nature of the encoded information.

Jacoby (1974, Experiment 3) presented a list of words in isolation in what he called the "looking-back procedure". Very simply, as every list item appears, the subject is required to say whether it is related either phonemically (Group 1) or semantically (Group 2) to any previous item in the list. In accord with Bregman's results, Jacoby found that the nature of the probe information had no effect on the recognition data: performance declined equally across 0 - 12 intervening items.

In each of the recent theoretical papers on depth of processing (Craik & Jacoby, 1975; Lockhart et al., 1976), it is suggested that the retrieval mode employed by subjects in each of these experiments was backward scanning. It is these data which lead Craik to the view that retrieval by backward scanning is unaffected by the nature of the information probed and in this way may be distinguished from retrieval by reconstruction. Numerous studies cited previously have shown large and consistent retention differences, with deep, semantic information showing superior retention to phonemic and structural information. Many other studies, often considered to be investigating semantic memory, have shown much higher retention levels for the semantic content of presented material in comparison to relatively low level structural information (e.g., Bransford & Franks, 1971; Sachs, 1967).

Other data reported by Jacoby (1974) and Craik and Jacoby (1975) lend substantial support to the efficacy of reconstruction as a retrieval process. These results are of particular interest because the same subjects provided evidence for both backward scanning and reconstruction on exactly the same test material. In Jacoby's experiment, subjects were presented a cued recall test subsequent to performing the looking-back task on the entire list. Despite finding no performance differences for phonemic and semantic information in the looking-back task, Jacoby observed highly significant differences in cued recall. Subjects making semantic judgments recalled more than double the number of words recalled by subjects making phonemic judgments (approximately 54% and 24% in his "n-back semantic" and "n-back acoustic" conditions respectively). Craik and Jacoby gave some of their subjects an unexpected free recall test after they had completed the encoding/recognition task. Again, while no performance differences occurred during recognition, semantically-encoded items yielded higher recall levels than phonemically-encoded items, and both were superior to structurally-encoded items.[11]

---

[11] Craik and Jacoby added a further comment on this data: "this extremely interesting result should be treated with some caution, as attempts to replicate the finding have yielded inconsistent and noisy data" (p. 186).

Craik argues that the handful of studies showing no differential forgetting of semantic and acoustic information reflect situations in which backward scanning was being employed by subjects. On the other hand, reconstruction was being used in those experimental conditions showing higher retention level for semantic over phonemic information.

There is unfortunately no statement by Craik indicating under what specific conditions either retrieval mode will be enacted. It is clear though that the amount of retrieval information is an important determinant. Thus when a substantial amount of retrieval information is made available (e.g., in a recognition test), backward scanning will be relatively useful over a large number of intervening items. However, as the retrieval information declines to a minimum (e.g., in a free recall test), reconstruction will be necessary even for relatively recent inputs to episodic memory. In this way, Craik and Jacoby found no differences in recognition performance with lags up to 24 items, thereby implicating the use of backward scanning. In contrast, many other studies have shown differential retention functions over very short lags and time intervals using free recall and cued recall procedures (see Baddeley, 1972, for a summary). In these situations, it is presumed that subjects are using the reconstruction process.

It should be evident from this discussion that the two retrieval modes are distinguished primarily in terms of the obtained data: when differential retention occurs, reconstruction is implicated; otherwise, backward scanning is suggested. In a very early section of this introduction, we noted that probably the most significant problem with all multi-store theories was their lack of explicitness in specifying what items are retrieved from STM. While many methods have been offered, none has gained consensus and, more importantly, none is clearly justified on the basis of any particular multi-store model. If the problem of identification is critical to those sorts of memory models, it is also the case with depth of processing: a clear theoretical delineation of the conditions inducing backward scanning and reconstruction must be elaborated. Otherwise predictions based on this feature of depth of processing will necessarily be vague and enigmatic.

For purposes of the present investigations, we have chosen to define reconstruction as occurring whenever a minimum of 90 sec intervenes between presentation of the last item of a long list and the beginning of the test trial. This definition is admittedly arbitrary to some extent but has been based on temporal parameters of many studies cited previously including Hyde and Jenkins (1969), Jacoby (1974), and Tzeng (1973). In each of these studies, "reconstruction effects" were found with intervals of approximately this length or shorter.

Comments

In the final section of this review, we wish to comment on the revised versions of the depth of processing model. A number of the issues raised in previous sections of the paper have been dealt with in these updated versions. These include the matters of retrieval processes, differential interference, and retention effects due to congruity. Each of these matters has been discussed at length in an earlier section. It was also pointed out in an earlier comment that depth of processing very clearly conforms to current memory theory in that it views memory traces as composites of attributes. This has been acknowledged very recently by Craik and Tulving (1975, p. 291). We will now turn our attention to issues which have not been dealt with and others which have developed in the revised versions of the theory.

One empirical finding which was problematic for the original model dealt with subjects who performed the characteristic orienting tasks but under intentional learning instructions. It has been found that even when subjects know they are to be tested later for their memory of the list items, performance on this test still reflects the typical depth effect (e.g., Craik & Tulving, 1975, Experiment 4; Mondani et al., 1973). It was argued that these subjects would be expected to process the items more elaborately after structural or phonemic orienting tasks in order that they might better retain them. Indeed, Craik and

Tulving (1975) suggest that "it is possible that further processing was carried out in the intentional condition, *after* the orienting task question was answered" (p. 278; emphasis is original). Logically, this should minimize the effects of the depth required by the orienting tasks if in fact depth of coding is the critical variable.

An alternative interpretation of these data is that subjects may in fact have stored deeper level information than that required by the orienting task but for one reason or another chosen not to use it. Rather, the very nature of the orienting task may have predisposed subjects to use specific information which they could remember from the earlier part of the experiment in order to reconstruct the targets. Were this the case, the observed depth effect would be predictable even under intentional learning instructions. It should be pointed out that Craik does not allude to this or any other interpretation of these otherwise perplexing and theoretically troublesome data. Unfortunately, there are apparently no data presently available which clearly demonstrate the use of such a strategy on the part of subjects. Therefore, this hypothesis remains conjectural.

Very closely tied to this issue is another important aspect of the depth effect. Craik and Tulving (1975) emphatically reasserted the hypothesis that semantic information is relatively more effective (in reconstruction) than lower level information: "a minimal semantic analysis is more beneficial for memory than an elaborate structural analysis" (p. 291). At the same time, however, they posit that a "minimal core encoding" will more or less automatically occur at all levels for very familiar, highly practised materials. Thus, common words may be expected to receive a modest analysis at structural, phonemic, and semantic levels even though the nature of the task may not require it. This minimal core encoding is basic in their interpretation of frequency effects in recall and recognition (see Lockhart et al., 1976, p. 95). If this is to remain as a critical feature of depth of processing, then it remains to account for the depressed retention levels of items encoded within the context of low level orienting questions. That is, if these items have the products of a minimal semantic analysis available in their respective memory traces, then why is this information not used at recall? Presumably, if it were, performance would be more in line with that found for items encoded with semantic questions (although of course the latter would have somewhat more elaborate semantic information available).

A possible explanation is the one suggested above. Specifically, the use of encoding questions during the study phase may induce their reactivation during the test trial. In this way, the orienting task procedure could be serving to predispose the typical depth effect. Thus, while it may be argued that the orienting task procedure provides some experimental control over the coding processes a subject exercises during the study trial (Craik & Lockhart, 1972,

p. 677), it may not provide the same control over the retrieval information activated during the test trial. This seems particularly true of free recall. Cued recall and some types of recognition tests are perhaps suggested.

If we accept this idea that subjects attempt to reconstruct using the original encoding questions, then we impose severe theoretical constraints on the utility of depth of processing. What we eventually arrive at is an a posteriori definition of the type of retrieval information the subject has used: if recall is poor, relatively low level information was used; if recall is good, deeper level information is implicated. Such a tautological approach will clearly be of little heuristic value in stimulating and interpreting future empirical work.

This problem brings us back to the matter of defining depth itself. We discussed this earlier and concluded that there must be some a priori basis for determining what types of analyses are deeper than others. This problem has not been rectified in the most recent statements of the theory. Indeed, it has been amplified such that not only must we find some basis for ordering the different levels, but now we must also determine which types of information represent deeper analyses within a given domain. Depth within and between domains is now a composite problem.

Our final comments deal with a few of the retrieval characteristics of the model. First, it is interesting to note in passing that Craik suggests a continuum of nominal retrieval cues. In most situations, it may be suggested that free recall conditions involve the least amount of retrieval information and recognition the maximum. Cued recall would normally fall in between. It seems, however, entirely possible that this ordering might be somewhat flexible as a function of many variables. For example, recognition performance can be decremented by manipulating the similarity of the distractor items relative to the targets (e.g., Anisfeld & Knapp, 1968; Underwood & Freund, 1968). Also, cued recall can be much superior to recognition depending on the nature of the recognition list itself (Watkins, 1974b). A detailed investigation of the conditions affecting performance under these various experimental manipulations would be of value, especially with the recent controversy over differences in recall and recognition (cf. Brown, 1976).

The three sources of retrieval suggested by Craik are consciousness, backward scanning, and reconstruction. We have already commented somewhat on consciousness and what other authors, primarily Mandler, have said about it. Since it does not seem to be of overwhelming importance to depth of processing, we do not wish to pursue the issue at length. However, we should comment briefly on the ambiguous use of the term. Specifically, our concern is with the actual capacity of consciousness. This is important to the extent that we will eventually need to separate items retrieved directly out of consciousness from those retrieved via backward scanning. In our earlier discussion, we talked about consciousness as if it had a capacity of one item, but this is not an entirely obvious conclusion based on its theoretical description. For example, Craik and Jacoby (1975) say that "as soon as a further perceptual event occurs, the last event is pushed out of mind [i.e., consciousness]" (p. 175). This would seem to suggest a capacity of just one item. But later, they say 'items "in primary memory" are still in conscious awareness' (p. 178), thereby implicating a somewhat larger capacity. This matter requires a clear theoretical statement from the authors.

Exactly the same criticism applies to the backward scanning retrieval process. As we have already discussed, the amount and nature of the retrieval information provided at test will determine in part whether backward scanning will be fruitful. We are simply left with the statement that backward scanning becomes "rapidly less efficient as the number of intervening items increases." While this may describe the obtained results, it does not lead us to a clear prediction as to how efficient this retrieval mode will be at any given point in time. More important, possibly, is the problem that we cannot predict when subjects will engage this process, in contrast to reconstruction, on any a priori basis. Lockhart et al. (1976, p. 89) have suggested that scanning may be activated when events are recent or experimental time constraints are imposed (e.g., paced testing). Apparently, then, strategies in part determine when backward scanning will be employed.

Another concern here is empirical. Craik suggests that this scanning process works in reverse order: that is, the last item input into episodic memory is the first item scanned. This seems to be a carryover from his earlier method for distinguishing items retrieved from STM and LTM respectively (Craik, 1968; see also Tulving & Colotla, 1970) since it uses the same reverse-order principle. This is confusing since there really are no data to support it. Murdock's (1974) comments vouch for this contention:

> As anyone who has ever done free-recall experiments knows, free recall is free in name only. The order in which the words are recalled is quite stereotyped. Generally subjects start with the last few items then, after running them off in serial order, jump back into the middle (or beginning) of the list, though still with a forward bias. (p. 202; emphasis added)

Much free recall research in our own lab confirms Murdock's view. It may be possible to modify backward scanning to allow for this, although it is not clear how this would be accomplished.

Since we cannot clearly isolate on an a priori basis items retrieved by backward scanning, we also, therefore, cannot unambiguously identify items retrieved via reconstruction. Apparently the only way in which these items can be identified with certainty is on the basis of their retention characteristics: that is, reconstruction has been used if the forgetting functions vary as a function of the type of encoded information. A somewhat more definitive theoretical separation of the products of these three retrieval processes is indicated.

The final matter we present here for the purposes of identifying an issue rather than to discuss it at length. Craik among others has used the term reconstruction to refer to a process by which the subjects attempt to reinstate a previous encoding context. But at a theoretical level, we can question whether the encoding context should be considered an actual part of the memory trace for an event. Thus, while we are clear on the effect of the encoding context on what is stored about an event (e.g., Light & Carter-Sobell, 1970), we do not know to what extent the context and the event are stored as a unified trace. If such a unified trace results, then perhaps the term "redintegration" might be a more appropriate description of the retrieval, notwithstanding Horowitz and Prytulak's (1969) arbitrary definition of redintegrative memory. The current distinction between the two terms seems to be that reconstruction occurs if the retrieval information is not actually a part of the desired memory trace whereas redintegration occurs if the cue is an intrinsic part of the trace. While this may seem a rather esoteric issue, it could potentially lead to some very interesting research which could further our understanding of the memory system.

Overview of the Research

A substantial amount of data has been presented in this review of depth of processing. Some of it has confirmed various aspects of the model while some has suggested where possible revisions may be necessary. In addition, though, a number of features of the model have yet to be put to experimental test. Two which will be of primary interest in the present investigations are the reconstructive process of retrieval and the concept of a minimal core encoding.

In regard to the reconstruction process, we should first note that the evidence for the other two retrieval modes is neither overwhelming nor unequivocal. For example, the primary evidence used in support of the backward scanning process derives from procedures similar to Jacoby's "looking-back" task. Unfortunately, though, the nature of the retrieval cues used in these types of experiments makes it rather difficult to distinguish the different types of retrieval processes on a procedural basis. Rather, the distinction, as noted previously, comes from the type of recall data obtained. Specifically, when there are no retention differences due to the nature of the encoded information, backward scanning is indicated; otherwise reconstruction is assumed. While we will not be expressly concerned with this issue in the present research, it will be important to determine some of the experimental parameters of this looking-back procedure in order to discover the conditions under which the depth effect ~~can~~ be evidenced.

A matter of present interest is whether data can be compiled showing that reconstruction is a viable retrieval process. While many authors have referred to reconstruction in one sense or another (e.g., Bartlett, 1932; Herriot, 1974; Russo & Wisher, 1976; Weingartner, Walker, Eich, & Murphy, 1976), none has defined exactly what is meant by the term. An operational definition is necessary. On the basis of evidence alluded to earlier, it was decided to impose a minimum of a 90 sec delay interval between study and test trials in each of the present experiments. Items retrieved subsequent to this delay were nominally classified as having been retrieved via reconstruction. It must be made very clear, though, that this definition is contributed solely by the present author for experimental purposes and perhaps would not be embraced by Craik and his colleagues.

The logic for the first two experiments was as follows. Retention of encoding information is assumed to be critical for the operation of reconstruction. In order to explicitly identify the encoding information, the orienting task procedure was employed. Thus, for reconstruction to occur, it would be necessary to show that subjects could in fact retain the encoding information for items correctly recalled. Thus, after performing standard recall tests, subjects were asked to write down the original encoding question for each item retrieved. If encoding information is retained for a very large percentage of recalled words, then the reconstruction principle receives correlational support; otherwise, a reassessment may be in order. This interpretation may be subject to certain constraints which are discussed later.

A second theoretical concept for which the extant experimental literature provides little or no confirming evidence is the minimal core representation. We have to date cited data which rather consistently revealed that much if not all incoming (verbal) information is automatically processed to a semantic level although no author has been able to show that the products of these analyses are in fact retained over the long term. This latter observation appears to be contrary to a belief in the continuity between perceptual and memorial processes and, more specifically, that semantically-processed information is relatively well retained in memory. These represent major features of the depth of processing model. Thus it could be suggested that a failure to obtain evidence of a semantic representation in memory for highly familiar words argues against the concept of a minimal core representation.

Two experiments were conducted in which semantic information was explicitly probed. It was hypothesized that words would be accessed by this testing procedure if in fact a functional core representation exists in memory. If evidence of a memorial trace at all levels does not obtain, then the theoretical utility of the concept must be brought into question.

Experiment 1

The first experiment was run for the explicit purpose of determining the degree to which reconstruction, as we have defined it, could serve as a viable interpretation of retention data. Is it possible to attribute all relatively "long-term" retention to this reconstruction process?

In this study, we will attempt to assess the degree to which subjects have retained the exact encoding information presented during the study trial. Six independent groups were presented a long list of items using the standard depth of processing paradigm. Three of the groups were later tested with free recall and the other three with 3-alternative forced-choice recognition. Subsequent to this test for the target items, all groups were asked to write down the specific encoding question associated with each item either free recalled or recognized. It was expected that this experiment would provide a rough estimate of the number of target items which could be retrieved using reconstruction. Presumably, if subjects are engaging in this form of retrieval, then they should have little difficulty in simply identifying retrieval information.

The results of a pilot study indicated that this procedure would be useful in providing data on the issue of retrieval of encoding information. While this preliminary study clearly suggested that encoding information was retained best of all for items paired with deeper level orienting questions, it was designed only to determine whether general encoding context was retained. In the present experiment, subjects were required to write the exact question associated with each word correctly recalled. Otherwise, this experiment was in many ways a replication of the pilot study, with only a few modifications related to experimental controls.

In summary, then, the first experiment was designed to provide some parametric data on the extensiveness of the hypothesized reconstruction process in which subjects generate the encoding information and subsequently use that to cue the recall of the target item.

General Method

As most of the experiments involved basically the same methodology, it will be convenient to outline the basic features of the experiments in detail here and simply note the particular variations on this method as we describe each experiment.

Subjects

Subjects were typically tested individually in an experimental session lasting 25 - 45 minutes. Order of assignment to conditions was according to a predetermined randomized block procedure.

Apparatus and Procedure

Subjects were shown a list of words on a T.V. screen situated approximately 4 m in front of them. Presentation of the words was monitored by a closed-circuit system in which a camera mounted at the back of the room was focused on the carriage of an electric typewriter. The words were typed on a long strip of paper inserted into the typewriter. A series of timers automatically controlled the advance of the carriage permitting reasonable accuracy in controlling stimulus onset and offset. An auditory transient from the typewriter signalled each event.

Subjects sat at a desk on which a response panel was mounted. Two seconds after the subject pressed the "READY" button, the word appeared on the screen. Simultaneously, a timer calibrated in milliseconds was activated. When the subject had made his decision, he pressed either the "YES" or "NO" button, each appropriately labelled to avoid confusion. These labels were reversed daily. A response stopped the timer and also turned on a small light so that E knew which response had been made. Presentation of the word was terminated after 3 sec although the response timer carried on if the subject had not made his decision by that time. It was a rare occurrence for an individual decision to take longer than 3 sec.

Subjects were read instructions which told them that they were participating in an experiment concerned with "how people perceive common words." They were told that they would be answering various types of questions about words appearing on the T.V. screen. Examples of each of these questions were included in the instructions. It was emphasized that both speed and accuracy were important in making their decisions.

The questions were typed individually on 5 x 9 in (12.6 x 22.8 cm) cards and mounted in a small 3-ring binder. Subjects were instructed to read the question carefully and to press the "READY" button only after they understood the question thoroughly. After they had responded to a question, they turned the card over and carried on to the next question. When the subject came to the end of the deck of cards, he was asked to fill out various administrative cards crediting him with experimental participation. This took approximately 90 - 120 sec. For simplicity, we shall refer to this entire procedure of answering questions as the "encoding task."

Lists and Encoding Questions

Typically, three levels of encoding questions were used. Structural questions were all of the form "Does it begin with the same letter as _____?" Phonemic questions asked "Does it rhyme with _____?" Semantic questions involved category information: "Is it a type of _____?" Exactly half of the questions asked at each level exhibited a congruous, and half an incongruous, relationship with the word appearing on the screen. These three levels by two types of relationships yield six "encoding conditions." While every subject in a given experiment saw exactly the same sequence of words, the type of encoding question was varied. Thus, for one third of the subjects in each condition, a given word was paired with a structural, phonemic, or semantic question. For each word, two of the three questions used were either congruously or incongruously related to it. Encoding conditions were presented in a randomized block design such that an equal number of each occurred in successive groups of 12 or 18 trials.

List items and the critical words used in the structural and phonemic questions were carefully selected so as to avoid, as much as possible, any overlapping characteristics with other words in the experiment. Of course, this was with the exception of the list word and the critical word in the associated encoding question. Also, rhyming items were selected so as to meet two criteria. First, no item in a pair of rhymes represented just a fragment of the other item: thus, "EACH-TEACH" was unacceptable. Also, approximately half of the critical rhyming items in the encoding questions were orthographically dissimilar to the list word (e.g., WAND-POND, CLUE-SHOE). Since each list used a minimum of 36 words and as many as 60, it was not possible to control the structurally-encoded items as stringently. Inevitably, the same initial letter had to be reused. This problem relates back to the issue of differential interference discussed earlier.

EXPERIMENT 1

Method

Subjects and Design

Fifteen subjects were assigned to each of six independent groups which are identified according to the nature of the learning instructions and the type of retention test. Groups 1 - 3 were tested with free recall and Groups 4 - 6 with 3-AFC recognition. Groups 1 and 4 were given incidental learning instructions: no mention was made of the later retention test in the initial instructions. Groups 2 and 5 were told that they would be tested on the target items as were Groups 3 and 6 who were also instructed that their memory for the specific encoding would also be tested. The actual type of test which these four intentional learning groups would receive (either free recall or recognition) was specified at the outset of the experiment.

Lists

A total of 54 encoding question-target pairs were used in this experiment. The first six, one from each encoding conditon, were practice items and were not included in any of the results to be reported. An equal number of subjects in each group was presented each of the three lists differing only in terms of the specific encoding associated with each of the targets.

All words used as distractors in the recognition test were as unrelated as possible to the targets with respect to both phonemic and categorical attributes. Exactly half of the items from each of the six encoding conditions were presented on this recognition test.

Procedure

After completing the orienting task and fulfilling the requirements of the delay interval, all groups were read appropriate test instructions. The free recall instructions were standard with subjects encouraged to set a lenient output criterion. The recognition groups were carefully instructed about the forced-choice procedure and the manner in which they were to rate their confidence in each recognition decision on a 4-point scale. A confidence judgment of 1 was to indicate a "guess" and 4 "certain". A card specifying the meaning of each of the values on this scale was left on the desk with the subject. No reference was made at this point to the subsequent retention test for encoding information. All groups were given a minimun of 3.5 min to perform this initial retention test and very few took greater than 5 min.

After this test had been completed, all groups were then asked to write the encoding which had been presented with each of the words that they had either free recalled or recognized. To simplify the subjects' task, number codes were used to refer to the type of question. The subjects had only to write down a number and the critical word or words from the encoding question. The instructions encouraged subjects to write an encoding question for each item identified as a target even if it meant guessing. This test was performed on the same page as the previous test.

In the final phase of the experiment, subjects were asked to write down any additional encoding questions which they might remember. This was done using exactly the same procedure as above for recalling encoding information. If any additional questions were written during this test phase, the subjects were subsequently asked to write the target word associated with each of these questions if it was available.

Results

Reaction times. The first data to be considered in all of the studies to be reported are reaction times (RT's). These data were collected from all subjects tested individually. Our comments on these data will be confined to reporting the effects of each of the major variables. Data suggesting that RT's can be manipulated independently of the depth effect were summarized previously where it was further argued that the specific types of questions selected by the experimenter would be important in determining the pattern of results in the RT data. Indeed, in the pilot study referred to in the brief introduction to Experiment 1, very long RT's were observed for structurally-encoded items (subjects were required to match a consonant-vowel pattern to the target word) compared to phonemic and semantic items (M = 2.101, 1.260, and 1.261 sec respectively). Despite this, the depth effect was clearly evident in its usual form. Jenkins (1974b; Walsh & Jenkins, 1973) has discussed this matter at length.

The instructional condition and test condition manipulations had no effect on RT's (both _F_'s < 1, _MSe_ = 3.488), as would be expected. There were, however, main effects of levels (_F_(2, 3948) = 147.41, _p_ < .01) and response type (_F_(1, 3948) = 90.10, _p_ < .01). The pooled within-subject error term for these and all other RT effects was 0.088. Structural judgments yielded lower RT's (_M_ = 0.971 sec) than phonemic (_M_ = 1.119 sec) or semantic judgments (_M_ = 1.149 sec), although there were no significant differences among these means according to a Duncan's multiple range test (_p_ > .05). Also, "yes" responses required less time overall than "no" responses (_M_ = 1.037 and 1.123 sec respectively). Both of these effects will be observed consistently in the later experiments since the same type of questions are used throughout.

Another main effect was observed in this experiment for replication (_F_(7, 3948) = 14.17, _p_ < .01). Essentially, this represents a practice effect which was reduced but obviously not eliminated by the initial six practice items. Reaction times in this and the later experiments typically asymptote after the third or fourth experimental trial in each encoding condition.

One problem which consistently arises in the analysis of RT's is that replications is completely confounded with items. Thus a number of spurious higher order interactions were obtained throughout this series of experiments which are probably attributable to item effects. As the successive studies were designed, new lists were constantly being constructed with an eye to these effects. However, we were not completely successful in excluding this particular problem from the data. Thus, while some interactions with replications appear to be meaningful and will therefore be referred to, many seem to be artifacts of the particular items used and will not be pursued in the discussion of the RT data.

One interaction which was reasonably consistent was response type x replications ($\underline{F}(7, 3948) = 4.09$, $\underline{p} < .01$) indicating a greater practice effect for "no" responses than for "yes" responses. Further, a second-order interaction of levels x response type x replications ($\underline{F}(14, 3948) = 1.83$, $\underline{p} < .05$) was obtained which supported the observation that the previous effect was somewhat greater for semantic questions than for phonemic or structural questions.

Another effect which occurs with regularity in the other studies to be reported is a levels x response type interaction ($\underline{F}(2, 3948) = 5.04$, $\underline{p} < .01$). In the present experiment, this reflects the fact that differences between "yes" and "no" responses were greater for structural questions than phonemic or semantic questions.

The only other significant effect found in the RT data of Experiment 1 was a small interaction of test condition x instruction condition x levels ($F(4, 3948) = 3.05$, $p < .05$). The only interpretation of this effect which seems plausible is that it represents a random effect due to subject differences. The power of this statistical test is such that some small differences might be expected to appear occasionally by chance. All other effects were nonsignificant (all p's $> .05$).

Recall. One of the most noticeable aspects of the recall data is the low overall performance level. Across the three instructional conditions, subjects recalled an average of just over 5 of the 48 experimental items, a mere 10.54%. While this may seem extremely low, it is not exceptional. Craik and Tulving (1975, Experiment 3) found overall recall of about 13% under comparable conditions. Craik (1973, Experiment 5) found recall of less than 10% for items presented once. Despite these apparent floor effects, significant differences due to certain variables have consistently been reported. It is within this empirical framework that we will consider the present data.

The recall data from Experiment 1 are presented in Table 1. The pooled within-subject error term for analysis of the recall data is 0.287. Analysis of variance revealed a main effect of levels ($F(2, 714) = 68.03$, $p < .01$), supporting the observation that semantic items were recalled best (22.12%) followed by phonemic (6.25%) and structural (3.31%) items. Similarly, we observed a main effect of response type ($F(1, 714) = 44.71$, $p < .01$) indicating that congruous items were better recalled than incongruous items (15.29% and 5.83% respectively). Both of these effects are quite typical.

Recall of Encoding Information. Recall was classified according to 1) retrieval of the target but no encoding information, 2) retrieval of the target and the level of coding (structural, phonemic, semantic), or 3) retrieval of the target and the complete encoding question. This response measure was included as a variable in the analysis and yielded a significant main effect ($F(2, 714) = 57.33$, $p < .01$). This was entirely due to more items being recalled with the complete encoding question (an average of 3.39 per subject) than target items only (0.84) or target items with just the levels information (0.83).

The main effect of instructional condition was not significant ($F < 1$, $MSe = 0.411$) nor did it interact with any other variables (all p's > .05). Two interactions which might have been predicted but did not materialize were instructional condition x levels and instructional condition x type of encoding information recalled. The latter effect might be expected if instructions to specifically retain the encoding information had an influence on this recall measure. The former effect was predicted on the basis of data cited earlier suggesting that lower level items benefit from intentional learning instructions whereas semantic items do not.

Each of the first order interactions of levels x response type, levels x type of information recalled, and response type x type of information recalled (all p's $< .01$) are constrained by the second order interaction of levels x response type x type of information recalled ($F(4, 714) = 39.58$, $p < .01$). As is evident from Table 1, this effect is primarily a function of the disproportionately large number of congruous semantic items which were recalled along with the complete encoding question. The magnitude of this effect is in part due to the differential recall of items across the six encoding conditions, although it is still very evident when the data are conditionalized on this factor. Indeed, for 93.60% (117 of 125) of the congruous semantic items recalled, the encoding question was also recalled. No analysis of variance was performed on this conditionalized data since it was felt that the small number of observations in some of the encoding conditions militated against stability.

With the exception of the three first-order interactions referred to above, no other effects approached significance (all p's $> .05$).

In the final test phase, subjects were asked to write down any additional encoding questions which they might be able to remember. Average output on this test was exactly one question per subject. For 71.11% of these additional encodings, subjects were unable to recall the target. Again, no analysis was performed due to the very small number of observations, although it appeared that the instructional manipulation did not affect performance on this test. On the initial free recall test, subjects output many more semantic items than lower level items. Despite this, recall of additional semantic encodings still represented 60.00% of the total on this later test.

Recognition. Overall level of target recognition was 69.35%, not corrected for guessing. Although a *d'* analysis might have been most appropriate for these data since confidence judgments were also collected, it was not possible to do this. As there were only four items tested from each of the six encoding conditions, too few data points were available to plot a reliable MOC curve. Consequently, analysis of variance was performed on the number of hits observed in each condition irrespective of confidence judgments. These latter data will be reported below.

The results are presented in Table 2. Hit rate did not vary as a function of the instructional condition ($F(2, 42) = 1.46$, $p > .05$). Informing subjects of the later retention test had no effect on overall recognition performance, as was the case for recall. Further, this variable did not interact with any other variables (all F's < 1). Levels again yielded a significant main effect ($F(2, 84) = 53.45$, $p < .01$) as did response type ($F(1, 42) = 11.00$, $p < .01$).[12]

---

[12] The within-subject error terms in this analysis could not be pooled due to heterogeneity of variance. The mean squares for the various error components ranged from 0.227 to 1.039.

These two results reflect respectively the standard depth effect and the superior retention of congruous compared to incongruous items.

The main effect of type of encoding information recalled ($F(2, 84) = 17.32$, $p < .01$) was due to the fact that, of all items recognized, level of encoding was recalled more often (40.20%) than the complete encoding (25.48%). A Duncan's test confirmed this difference ($p < .01$). Also, recall of no encoding information occurred more frequently than recall of the entire encoding question according to the Duncan's test ($p = .05$). The apparent superior retention of encoding level (compared to retention of no encoding information or retention of the complete encoding question) may be an artifact of guessing strategies. Subjects were instructed to write down as much of the encoding question as possible, even if it meant guessing. Thus, it would be expected that subjects would have at least a one in three chance of correctly guessing the level of the encoding question even if they had no idea what it was originally. The guessing data to be presented below indicate that a guessing rate of one third is conservative, at least for structural and phonemic items.

As in the recall data, type of information recalled interacted with levels and response type ($F(4, 168) = 36.21$, $p < .01$). This is mainly a function of the disproportionately large number of congruous semantic items for which the complete encoding was recalled. Looking at these data in terms of proportions of all items recognized from within a given encoding condition does not alter the interpretation of these data: in no encoding condition did retention of the complete question exceed 16% of all items correctly recognized except for congruous semantic items where this value was 87.06%. The interactions of levels x type of information recalled ($\underline{F}(4, 168) = 47.33$, $\underline{p} < .01$) and response type x type of information recalled ($\underline{F}(2, 84) = 97.36$, $\underline{p} < .01$) are qualified by this higher order interaction of the three variables. The levels x response type interaction ($\underline{F}(2, 84) = 2.40$, $\underline{p} < .10$) approached but did not attain significance. It too is qualified by the higher order interaction.

Subjects in the recognition conditions generated an average of just 1.09 additional encodings in the final test phase of the experiment. The target was also recalled for 63.26% of these additional encodings, primarily due to the large number of congruent semantic items accessed through this information (48.98% of all encodings generated were congruous semantic for which the target was also recalled). The instructional manipulation again had no apparent effect on this test.

Confidence Judgments. Due to missing data for a number of subjects, these data could not be analyzed according to normal statistical procedures. However, the data will be summarized here in descriptive form. First, consistent with all other data reported, the instructional manipulation had no consistent influence on confidence ratings, although mean confidence rating was slightly but unremarkably lower for Group 4, the incidental learning condition.

Mean confidence judgments for hits reflected the depth effect entirely: confidence in recognition decisions increased directly with depth and was consistently greater for congruous than for incongruous items. These values are presented in Table 3 along with the number of cases on which each mean is based.

Also presented in Table 3 are the confidence judgments for false alarms. Of course, as would be expected, there are no differences in these values since encoding condition is a pseudovariable if the subject selects a distractor item.

Recall of Encoding Information. On the initial test trial, subjects were asked to write the level of the encoding question plus the critical words of the question beside each word either recalled or recognized. Those items for which correct retrieval occurred but the level of the encoding question was incorrectly specified provide interesting data on guessing strategies. The low overall level of free recall did not permit an extensive analysis of the guessing data for level of encoding, although consistent trends were still evident. If the specified level of encoding was incorrect, subjects chose the lower level of encoding for 73.33% (22 of 30) of the items. Thus, if an item was encoded phonemically, subjects were more inclined to call it a structural item than a semantic item when they did not correctly identify it as phonemic. Of course, we do not know how often the correct level was guessed by our subjects although incorrect guessing occurred for only 13.24% of all items recalled.

The recognition groups provide two sets of data for our consideration. First, for those items which subjects correctly recognized but incorrectly guessed the level of encoding, the lower level of encoding was selected for 72.24% (187 of 251) of the items. These data are comparable to that reported for the free recall groups, and again did not vary across instructional conditions. The second set of data is that for false alarms, items incorrectly identified as targets. Subjects can of course select any of the three levels of encoding for these items. However, they still tend to specify the lower encoding levels for these items: 53.26% were assigned to the structural level, 35.91% to phonemic, and only 10.84% to semantic.

These data seem to reveal a very clear guessing strategy on the part of subjects when they do not remember the level of encoding associated with a given item. Semantic encodings are apparently much more memorable than the lower level encodings and thus, when the subject is in doubt, he will rarely say that an item was encoded semantically; rather, he will select a lower encoding level. This result must be considered in an appropriate correction for the data reported earlier showing retention of encoding level following the forced-choice recognition test. For example, for all the structural items correctly recognized, encoding level was "recalled" for 64.70% (121 of 187) of them. The corresponding value for phonemic items is 54.42% (123 of 226). In both cases, a biased guessing rate must be taken into consideration when interpreting these data in terms of retention of encoding information. Indeed, the finding that the complete encoding question was recalled for only 1.07% (2 of 187) of the structural items and 8.85% (20 of 226) of the phonemic items correctly recognized argues against the use of this information at the time of testing and thus against the use of a retrieval process such as reconstruction for these items. Subjects cannot effect retrieval via reconstruction if they cannot remember the encoding information, according to the present operationalization of the term.

Discussion

In this first experiment, we had hoped to provide some evidence on the retention of encoding information presented during the study trial. The availability of this information would seem to define the extent to which reconstruction, as we have operationalized it, can serve as a viable source of retrieval from relatively long-term memory.

The results of this experiment would seem to suggest two conclusions. First, the retention of encoding information for words elaboratively processed at the semantic level was quite substantial, particularly when the encoding question and the target word were congruously related. It seems entirely possible, then, that this information could be utilized by the subject during the test trial in order to reconstruct the target word. Secondly, there was little evidence for the retention of encoding information for items processed at the lower structural and phonemic levels of analysis. For these items, then, reconstruction could not be considered a viable mode of retrieval. While free recall levels were insufficient to permit a good test of this hypothesis, recognition performance was well above chance. Still it was the rare occasion that subjects were able to identify the encoding question associated with these target items. The obvious interpretation of this finding is that even when such items are retrieved, it is probably not through a process such as reconstruction. Exactly what type of process is involved in the activation of the memory traces for these items cannot be discerned from the present research.

Additional evidence for these conclusions is available from the confidence judgments supplied by subjects performing the recognition task. When subjects were able to identify the encoding question, confidence judgments were extremely high. For items congruously encoded at the semantic level, 97.30% (144 of 148) were assigned the highest confidence rank when the encoding information was correctly recalled; the comparable value for the incongruous semantic items was 100% (21 of 21). For items encoded at lower levels, there were of course many fewer instances overall. However, 85% (17 of 20) of the congruous phonemic items were given the highest confidence rating when the complete encoding was correctly identified; both (2, of 2) congruous structural items of this type were also assigned the highest rank. This strong correlation between the retention of encoding information and the confidence subjects have in their recognition decisions provides additional support, albeit peripheral and possibly equivocal, for the suggestion that encoding information may be important in retrieval.

While these aspects of the present data and that of the pilot study reported earlier provide correlational evidence for the possible existence of a reconstruction process as we have described it, it must be noted that some potential problems exist in such an interpretation. First, subjects may in fact have provided their own idiosyncratic encoding of the targets while they were being presented. During the test trial, they could make use of this other retrieval information in order to effect recall or recognition. This would seem most likely to occur in the groups specifically forewarned of the impending test. It is not clear to what extent this represents a significant problem in the present research. Of course, this experimental paradigm was designed to avoid exactly this sort of "confound," as we discussed earlier. It might also be noted that this additional encoding of items and the use of that information at retrieval is not necessarily in contradiction to the earlier suggestion that the nature of the encoding task may predispose the later reactivation of encoding information: the potency of the orienting task still exists and it may not be reasonable to imply a functional difference between encodings supplied by the experimenter and those generated by the subject.

This potential problem of subjects using their own encodings during retrieval would operate to underestimate the actual amount of reconstruction which is occurring since our test procedures are not designed to identify this form of retrieval information.

A second factor needs to be considered in the interpretation of the results of this first experiment. It seems entirely possible that subjects could be recalling or recognizing target items by some procedure other than reconstruction, although Craik does not offer any alternatives for relatively long-term retrieval. It is conceivable that subjects could be accessing stored items by some means other than reconstruction and only then generating the encoding question to fulfill the task demands. Thus, the fact that subjects are capable of specifying the encoding does not prove that this information is actually being used to effect recall or recognition of the targets. Of course, we do know that subjects can use this information when it is made available in order to improve retention: many studies have shown cued recall to be superior, normally, to free recall (e.g., Craik & Tulving, 1975, Experiment 7). However, increased recall under these experimental conditions does not allow us to conclude that the same information is used by the subject when it is not explicitly provided by the experimenter. It does not seem reasonable to argue the contrary position on the basis of such data since we must clearly allow for the fact that subjects tested under free recall conditions would probably not be able to access all the encoding questions which are made available to subjects performing cued recall.

The possibility of subjects retrieving items by some procedure other than reconstruction would suggest that our measure of the latter process might be overestimated. The extent of this overestimate is uncertain. Items recalled or recognized but for which no encoding information is retained might give us a rough idea of the degree to which other retrieval processes are used, although it is also possible that these items were retrieved by the subject reconstructing some idiosyncratic encoding, as was suggested above.

Finally, we do not know why semantic information serves as a better retrieval cue than lower level cues, although perhaps the matter of differential interference is relevant here. More important in terms of a reconstruction position is the question of why semantic encoding information is itself retained better than lower level information.[13] Craik does not offer more than a definitional statement in this regard. Perhaps the views of other authors (e.g., Restle, 1974) could be incorporated to provide some insights on this particular issue.

Finally, and a point which we wish to emphasize here, is the observation that a substantial number of target items were accurately retrieved without any evidence of the encoding information being accessible. While the empirical support for this conclusion derives mainly from the subjects performing the recognition test, there is also some limited evidence in the free recall data. However, again the low performance levels exhibited by the free recall subjects militate against any substantial theoretical remarks.

Subjects given the recognition test were often able to identify targets with a high degree of confidence even though they could not recall any of the associated encoding information. For example, of all incongruous semantic items correctly recognized and assigned the highest confidence rating, 45.79% (49 of 107) gave no evidence of any encoding information being retained. This represents 33.79% (49 of 145) of all incongruous semantic targets identified on the recognition test. Another good example comes from the incongruous phonemic items correctly recognized. For those items assigned the highest confidence rating, no encoding

---

[13] This question is interesting not only in terms of reconstruction. If one prefers the view that subjects activate the encoding information only after the target has been retrieved and perhaps only because of task demands, the differential retention of semantic information is still a legitimate concern.

information was recalled for 47.06% (24 of 51) of them. This represents 21.05% (24 of 114) of all incongruous phonemic items correctly recognized. Of course, the former figure is probably a slight underestimate since subjects were essentially forced to guess the encoding level thus undoubtedly inflating the number of items for which encoding level was nominally scored as correctly recalled. There was comparable data of this sort from each of the other lower level encoding conditions suggesting that items were often retained quite accurately without concomitant retention of the encoding information. The implication of these findings is that reconstruction is probably not the only retrieval mode for relatively long-term traces. Rather, some other process which does not require the retention of encoding information seems necessary particularly in terms of recall of items encoded at lower levels of processing. As mentioned previously, there is no evidence from the present research to indicate the nature of such a process.

At this point, then, we can summarize the first experiment by saying that some evidence for the existence of a reconstruction retrieval process has been provided, although this was largely restricted to items coded at relatively deep levels in the processing system. As a note of caution, though, these data are subject to alternate interpretations. Further, there is a clear suggestion that a substantial number of target items may be retained without their respective encoding information being accessible to later recall. Thus reconstruction may not be the only long-term retrieval process available to subjects.

It must be emphasized here that the present use of the term reconstruction may not reflect exactly that of Craik and his colleagues. We have endeavoured to describe this process in such a manner that it is as consonant as possible with the theoretical account provided in recent descriptions of the depth of processing model. However, these descriptions were comparatively vague as to the exact nature of process and thus it was necessary to operationalize reconstruction in order that the present investigation could proceed. It is in this experimental definition of reconstruction that we may have deviated from the original meaning intended by Craik. Thus, it must be acknowledged that the conclusions outlined above are restricted to the present use of reconstruction.

In this first experiment, we have shown that subjects are able to recall specific encoding information during the test trial. This was seen as necessary in order to make plausible the concept of a reconstruction mode of retrieval. What we have not done, however, is to demonstrate that this encoding information is actually used by the subject to access target items. It would be ideal, then, to present data now which documents the use of this process under free recall and recognition testing conditions. Unfortunately, given presently available procedures, a clear answer to this question is not evident. We are left, then, in the uncomfortable position of having only correlational data as our empirical support for the hypothesized process of reconstruction. The need for definitive evidence of this sort implores additional research.

Experiment 2

In the first study, we presented subjects with the standard orienting task procedure and found that retention of encoding information was highly correlated with target memorability. The inability to access the original encoding questions for lower level items suggested that perhaps reconstruction may not be a viable means of retrieving these items. Further, the relatively weak "relational encoding" (cf. Schulman, 1974) between target and question may magnify the inadequacy of these questions to serve as retrieval cues in those few instances in which they may be recalled.

One excellent means of dealing with the latter problem is to equate the strength of the relationship between the target and the encoding information at various levels. If some or all of the depth effect is a function of these poor "associative bonds" for low level items in contrast to items coded at a deeper semantic level, then perhaps the magnitude of the effect can be reduced either by increasing the associative strength between lower level encodings and targets or by decreasing the strength between deeper semantic encodings and targets. In a study referred to previously, Nelson et al. (1974) controlled for associative strength of phonemic and semantic cues through the use of normative data. Subjects saw lists of word pairs in which the target word was either phonemically- or semantically-related to the cue word presented simultaneously. Following presentation of 24 such pairs at a 3 sec rate, a cued recall test was given using the original study cues. While no direct empirical comparison is made with other findings, it is clear that the differential effectiveness of phonemic and semantic cues is substantially reduced. When the cues were all relatively strong, semantic targets were retrieved with approximately 89% accuracy and phonemic targets with 80% accuracy. These data suggest that the depth effect may be in part attributable to the overall commensurability of the target and its encoding, although we must be careful to note that semantic encodings were still somewhat more effective cues than were phonemic encodings.

It seems possible, then, that the magnitude of the depth effect which is typically obtained may be due in part to uncontrolled differences in the strength of the associative bonds between targets and the experimenter-supplied encodings at the various levels. Under normal learning conditions, subjects are not usually provided explicit encoding information. Thus, it seems important to assess the effect of depth of processing when the subject provides his own idiosyncratic encodings at different levels.

There has been a limited number of experiments reported in which subjects overtly supplied their own encodings.

Johnston and Jenkins (1971) performed one such study which we previously had occasion to discuss in detail. Subjects in one condition generated a rhyme for the target item. In a second condition, subjects generated an adjective associated with the target if the latter was a noun, or vice versa. The results showed a very clear depth effect: semantically-encoded items were recalled much better than phonemically-encoded items. Elias and Perfetti (1973) had independent groups generate synonyms, semantic associates, or rhymes to the list of target words. Using a yes/no recognition test which factorially combined encoding conditions with type of test distractors (synonyms, associates, rhymes), they observed higher confidence ratings for targets encoded in either semantic condition compared to the phonemic group and a control group simply instructed to learn the list. Further, the phonemic group gave higher recognition ratings to distractor items than any of the other groups. These results again reflect a typical depth effect, although in a somewhat different experimental setting.

In neither of these studies, though, was a comparison made between items for which encodings were supplied by the experimenter and items for which the subjects generated their own encodings. Thus, it is not possible to determine whether the depth effect is mitigated in the latter condition. Additionally, if such differences did occur, there is no means of assessing whether it was due to increasing the probability with which the encodings were retrieved (and thus made available for reconstruction) or if it was a function of increasing the cueing potential of the encodings once they had been retrieved. In the next study to be reported, experimental manipulations were introduced which allow these various comparisons to be made.

Two groups of subjects were presented with the standard orienting task procedure. Four additional groups were run, of which two generated their own targets from the encodings supplied to them, and two which performed the complementary task of generating encodings for the targets which were supplied. For the latter groups, subjects were required to generate an encoding at one of three specified levels. All groups performed this initial task under incidental learning instructions and were subsequently required to recall the targets under either free recall or cued recall test conditions. All free recall subjects were then asked to write down the exact encoding question which they had been given or they themselves had generated for each target item recalled. With these various manipulations, we should be able to determine whether subjects' own encodings lead to better retention and whether any retention differences can be specifically attributed to the memorability or the cueing potential of the encodings.

Method

Subjects and Design

Twenty subjects were assigned to each of six independent groups. Groups 1 and 2 performed the standard orienting task, with each subject being tested individually. Groups 3 and 4 were required to generate target items which were congruously related to the encoding questions given to them. Groups 5 and 6 were given the target items and required to generate encodings at a specified level. Subjects in these four conditions were run in groups of 2-5 for the purpose of experimental expediency. Level of encoding (structural, phonemic, semantic) was a within-subject manipulation as usual. All learning was incidental.

Lists

In this experiment, 60 question-target pairs were used. No practice items were presented in any of the conditions. Groups 1 and 2 used two randomizations of the target items with each item being assigned to a different encoding condition within the two orders. For these two groups, exactly half of the encodings at each level were congruous and half incongruous.

The encoding questions used on Groups 3 and 4 and the targets used in Groups 5 and 6 were the same as those presented to the first two groups. Half the subjects in all groups saw each randomization. While it would have been ideal to equate the number of congruous and incongruous encoding question-target pairs across all groups, it seemed impractical to ask Groups 3-6 to generate incongruous pairs. Consequently, subjects in these groups were asked to generate congruous relationships for all items.

Procedure

Study instructions for Groups 1 and 2 were entirely typical according to the General Method. Groups 3-6 were given two books, one labelled "Questions", the other "Answers". For Groups 3 and 4, the Question book contained ten pages, with each page having six complete encoding questions. The Answer book also contained ten pages, but each page simply had a series of six blank lines. For Groups 5 and 6, the Question book was identical except that only partial encoding questions were written. Thus, if a structural encoding was required, the sentence in the Question book would read "It begins with the same letter as _____" with a blank included for the subject to complete the sentence. The phonemic and semantic questions were respectively "It rhymes with _____" and "It is a type of _____". The Answer book for these two groups had six target words written on each page. All groups were given explicit examples of how they were to complete the task. Subjects in the latter four groups were all told to write down targets or encodings, as the case may be, such that they did not repeat items which they had written earlier or had seen in either book previously. Otherwise, they should write the first thing that came to mind which would be correct in the specific circumstances. Subjects worked at their own pace and turned over both books as soon as they had completed this task. All study conditions required approximately the same amount of time, usually from 10 to 15 minutes.

After all subjects in a group had completed the orienting task, a minimum of a 90 sec delay was imposed. Subsequently, all subjects were read the test instructions. Groups 1 and 2 were read the standard free recall and cued recall instructions respectively. For Group 2, six items from each of the six encoding conditions were probed with the previous encoding questions. The test trial for Groups 3 and 4 was identical to that of the first two groups, Group 4 being cued with exactly the same set of encodings as Group 2. Group 5 received the standard free recall task. During the delay interval, a second experimenter selected six pages of generated encodings from the Question book of subjects in Group 6 and rearranged them in a predetermined random order. The pages selected and the randomization of these pages was changed for each experimental session of subjects in this group. Since the cues presented to Groups 2 and 4 were completely randomized, the exact items tested in Group 6 were not identical to those used in the other two groups. The second experimenter was only necessary for the smooth running of Group 6 and thus was not present for any of the other groups. This second experimenter played no role in the session except to select and randomize the test pages; otherwise, he sat inconspicuously at the back of the room.

After the three ~~free~~ recall groups had completed the unpaced test, they were asked to write down the exact encoding question associated with each target recalled. As in Experiment 1, subjects coded their response and simply filled in the critical word from the question. The instructions encouraged subjects to write an encoding question for each item recalled, guessing if necessary. Since the first study had evidenced such poor performance on the final test in which subjects were asked to attempt recall of additional encodings, this was not required in the present experiment.

Results

Reaction times. Test condition had no significant main effect on RT's ($F(1, 38) = 3.49$, $MSe = 4.98$, $p < .01$), nor did this variable interact with any other manipulations (all p's $> .05$). The pooled estimate of the error used in all subsequent tests was 0.070. There were the standard main effects due to levels ($F(2, 2242) = 63.67$, $p < .01$) and replications ($F(9, 2242) = 24.74$, $p < .01$). Responses to structural questions were faster than to phonemic questions and both were faster than to semantic questions (M = 0.917, 1.001, and 1.066 sec respectively). Reaction times attained asymptote after approximately four complete replications. The failure to find a main effect of response type ($F < 1$) seems largely due to phonemic questions: "no" responses were actually faster than "yes" responses. Structural and semantic "no" responses took longer than the "yes" responses. This differential effect of response type across levels is verified by the significant first order interaction of the two variables ($F(2, 2242) = 8.17$, $p < .01$). All other interactions (with replications) appear to be due to item effects and will not be considered.

Recall: *Groups 1 and 2*. The recall data for the first two groups are presented separately in Table 4. Analysis of variance revealed that the cued recall group output more targets than the free recall group ($F(1, 38) = 9.11$, $MSe = 0.53$, $p < .01$). The main effects of levels ($F(2, 190) = 191.72$, $p < .01$) and response type ($F(1, 190) = 192.32$, $p < .01$) were entirely typical. The error term for these and subsequent effects was 0.597.

The interpretation of each of the first order interactions of test condition x levels ($F(2, 190) = 9.43$, $p < .01$), test condition x response type ($F(1, 190) = 34.20$, $p < .01$), and levels x response type ($F(2, 190) = 96.32$, $p < .01$) is constrained by the three-way interaction of test condition x levels x response type ($F(2, 190) = 10.68$, $p < .01$). This higher order interaction is almost entirely a function of congruous semantic items being retrieved better when the original encoding question was explicitly re-presented at test compared to when it was not. In other words, cueing was differentially effective for these particular items.

Since Group 1 is directly comparable with the free recall groups of the previous experiment, a separate analysis was performed on these data. This analysis varied from that reported immediately above in that type of encoding information recalled was included as a variable. Briefly, it will simply be noted that all main effects and interactions were significant (*MSe* = 0.25, all *p*'s < .01). Each effect mirrors that found in the earlier study. The best summary of these data can be drawn from the significant second order interaction of levels x response type x type of information recalled: a disproportionate number of congruent semantic items were recalled along with the complete encoding question in contrast to all other items.

Recall: All Groups. The four groups of subjects required to generate either encodings or targets were specifically advised that only congruous relationships should be used during the study trial. Thus, the present analysis compares all groups on only congruous items: response type does not enter in as a variable. Similarly, type of encoding information recalled is only relevant for the free recall groups and will be reported separately.

Since the groups differed in terms of how many items were actually tested, the analysis was performed on the proportion of items recalled as a function of maximum possible recall. Thus, Group 1 could recall a maximum of 10 items at each of the three levels (congruous only) whereas Groups 3 and 5 could output a maximum of 20 items. Similarly, Group 2 was probed for 6 items and Groups 4 and 6 for 12 items at each level. The results are shown in Table 5. Main effects were obtained for study condition (*F*(2, 114) = 128.06, *p* < .01) and test condition (*F*(1, 114) =

391.01, $p < .01$). These variables also interacted ($F(2, 114) = 13.69$, $p < .01$) indicating that the advantage of cued recall over free recall was maximal for the group generating their own targets and least for the group performing the standard orienting task. The error term for each of these effects was 253.82. A Duncan's multiple range test revealed that Groups 1 and 5 did not differ significantly from each other, nor did Groups 2 and 3. All other group differences were significant (all p's $< .01$).

Also observed was a main effect of levels ($F(2, 228) = 350.90$, $p < .01$): semantic items were retrieved better than phonemic and structural items. The error term for this and each subsequent effect was 165.42. The interpretation of the first order interactions of study condition x levels ($F(4, 228) = 17.12$, $p < .01$) and test condition x levels ($F(2, 228) = 81.58$, $p < .01$) was constrained by the second order interaction of study condition x test condition x levels ($F(4, 228) = 17.42$, $p < .01$). To facilitate the understanding of this interaction, it is plotted in Figure 1. Although the cued recall groups were able to output many more semantic items than the free recall groups, the differences due to study condition were consistent (possibly due in part to ceiling effects). However, at the lower levels of coding, free recall differences due to study condition were magnified under cued recall.

It will be evident from Figure 1 that items encoded structurally were actually recalled slightly better by Group

3 (generate targets, free recall) than items encoded phonemically. This finding should be treated with some caution. While it suggests perhaps that some of the effects of depth are mitigated when subjects generate their own targets, it was entirely possible for subjects to write a target which was both structurally and semantically related to the critical word in the encoding question. For example, in one instance, a subject was asked to write a word which began with the same letter as "CITY". His response was "COUNTRY". Apparently, then, the encoding question was nominally structural but the subject output a response which had a clear semantic relationship to the critical word in the encoding question. Other such examples were evident. However, it was not possible to quantify this problem since many of the potential examples were rather marginal and it is difficult to assess the influence of idiosyncratic associations on these data. A related problem was the generation of items which were highly similar to the critical word. For example, when asked for a word beginning with the same letter as "ESSAY" or "BEARD", subjects occasionally responded with "EASY" and "BREAD" respectively. These types of relationships were explicitly avoided in lists which the experimenter selected.

In considering these observations, two points should be kept in mind. First while generation of these types of items contaminates the pure structural nature of the relationship between encoding question and target, it may mean that subjects try to avoid purely structural encodings when given the freedom to do so. The basic effect of levels of encoding attests to the value of such a strategy. Secondly, when subjects encode an item with some additional higher level code, it is necessary that this additional information be accessed by subjects at the time of retrieval. Thus if the structural encoding "It begins with the same letter as CITY" is given on the test trial (or is activated by the subject himself), it will be relatively insufficient in itself as a retrieval cue for "COUNTRY" unless the subject also remembers that the target was semantically related to the critical word in the question (cf. Baddeley, 1972).

In summary, it appears that in some way subjects' own encodings are more beneficial at recall than are encodings provided by the experimenter. This is particularly evident under cued recall conditions and for lower level encodings, although again ceiling effects may have mitigated differences for items coded semantically. In order to determine whether the encodings themselves are more accessible, we will consider the accuracy with which the free recall groups were able to recall the encoding information.

Group 1 had only 10 opportunities to recall items in each of the congruent encoding conditions whereas Groups 3 and 5 had 20 such opportunities. Thus, in order to overcome biases which would occur in the analysis of raw scores, recall was translated into proportions of total recall opportunities. Study condition had a significant main effect on performance ($F(2, 57) = 49.49$, $MSe = 49.64$, $p < .01$). A Duncans' test confirmed the observation that Group 3 recalled more items than Groups 1 and 5 ($p < .01$). All other effects are based on a pooled error term of 45.31. Main effects were obtained for levels ($F(2, 456) = 70.41$, $p < .01$) and type of encoding information recalled ($F(2, 456) = 241.75$, $p < .01$). A greater percentage of semantic items were recalled (37.58%) than phonemic (15.83%) or structural (15.58%) items. Also, more items were recalled along with the complete encoding question (16.53%) than with just the level information (4.64%) or no encoding information at all (1.83%).

The interactions of study condition x type of information recalled ($F(4, 456) = 14.38$, $p < .01$), levels x type of information recalled ($F(4, 456) = 122.22$, $p < .01$), and the small interaction of study condition x levels ($F(4, 456) = 3.04$, $p < .05$) are qualified by the higher order interaction of study condition x levels x type of information recalled ($F(8, 456) = 4.68$, $p < .01$). This interaction is plotted in Figure 2. Basically it shows that whenever Groups 1 and 5 recalled a target word, they almost invariably recalled the complete encoding question associated with it. Group 3, on the other hand, often recalled a target word either by itself or with just the level information. Groups 1 and 5 retrieved the complete encoding for 83.11% and 92.43% of all targets respectively whereas Group 3 retrieved this information for only 55.61%

of all recalled items. This interaction is also in part a function of the much better retention of lower level items by Group 3, as may be seen in Figure 2.

Discussion

The intent of Experiment 2 was to determine whether the depth effect is mitigated when subjects generate their own targets or encodings in contrast to the standard orienting task procedure. More specifically, is the relatively poor retention typically observed in the latter conditions in whole or in part a function of subjects not being able to access encoding information (and consequently having no basis for reconstruction) particularly for lower level items or is it that the relational encodings themselves are rather inefficient for accessing target items?

Considering the cued recall groups, it is apparent from Figure 1 that the depth effect is certainly not reduced by having subjects generate their own targets or encodings. This conclusion holds for the structural and phonemic levels of coding. It is not possible to determine if the magnitude of the effect is reduced at the semantic level since ceiling effects were evidenced. Overall performance levels were increased under what might be classified as typical learning conditions compared to the standard orienting task procedure in which all targets and encodings are provided by the experimenter. However, this improvement in performance is not differential across levels of coding: relational encodings at the structural level are less well retained than phonemic encodings and both are less efficient than semantic encodings.

With regard to the free recall groups, we can again conclude that overall performance is much improved when subjects are given the opportunity to establish some part of the perceptual trace. The small interaction between study condition and levels suggested that perhaps the depth effect was reduced somewhat under the latter conditions. Group 5 recalled more items than Group 1 although this difference was entirely attributable to items for which the complete encoding information was retained. This is true for all levels of coding and provides good support for the concept of reconstruction. Indeed, few items were recalled by either group for which this information was not available. Apparently, subjects' own encodings are more accessible during the free recall trial than encodings supplied by the experimenter.

Retention of targets by Group 3 was clearly superior at all levels to that of the other two groups. A comparison between Group 3 and the other free recall groups is particularly interesting. While this group recalled more items overall, it is clear that retention of encoding information was not as good as it was in the other conditions. Of all items recalled by this group, the complete encoding information was retrieved for only 55.61% of the targets. This compares with 83.33% and 91.70% for

Groups 1 and 5 respectively. It might be suggested then that it is the specific information which subjects generate (targets for Group 3, encodings for Group 5) that benefits most at recall. Despite this, depth of processing is still a critical factor determining relative retention levels of different items.

The data of Group 5 provide strong support for the suggestion that subjects do in fact use encoding information at retrieval since targets were rarely recalled without the encoding context. It is this group which most closely resembles the standard learning experiment in that subjects supply their own encodings. Thus, the strong correlation between retention of targets and encoding information reinforces the contention that the latter is critical for accessing of targets. With Group 5 providing many more observations at the lower levels of coding, it is also possible now to generalize this statement to more than just semantically-coded items, as was concluded in the discussion of the first experiment. Finally, target recall by this group compares favourably with the effects found in the previously cited studies in which subjects generated their own encodings at various levels of analysis.

Experiment 3

One of the recent changes to depth of processing has been the identification of a minimal core encoding or representation. As described in the general Introduction, this refers to the idea that, with highly practised stimuli, a minimal analysis at all levels of processing is automatic and cannot be arbitrarily terminated by the nature of the orienting task. Since processing of information defines the establishment of a memory trace, it follows that a minimal memory representation of the input will be developed at all levels. Any additional processing carried out in order to meet the demands of the orienting task will of course result in an elaboration of the information stored from the specified level.

The intent of the next two experiments is to provide empirical tests for the existence of the minimal core representation. The logic of both experimental designs to be used is as follows. We allow that, regardless of the nature of the orienting task, a minimal semantic analysis will occur when subjects are presented a common high-frequency word. We further allow that "a minimal semantic analysis is more beneficial for memory than an elaborate structural analysis" (Craik & Tulving, 1975, p. 291). On this basis, it might be predicted that inducing subjects into using stored semantic information during the test trial would lead to an overall improvement in performance on the memory test. This would be particularly true for items coded at relatively lower levels in the system (structural, phonemic) since it is these items for which retention is generally very poor. This follows if in fact semantic information is a more effective retrieval cue than lower level cues. The implicit assumption here is that when subjects perform a lower level orienting task, it is that specific encoding information which serves as the retrieval cue normally. Thus, despite theoretically having available more efficient retrieval information, subjects choose, for whatever reason, to use that which derives directly from the orienting task. In the next two experiments, the test trial is designed in such a way that deeper semantic information is made particularly salient. If a minimal core representation exists in the memory trace for these items, we would expect to observe its influence on the retention test in the manner described above.

In Experiment 3, one group of subjects was presented retrieval cues which were identical to the original encoding questions. Another group was given cues which were all semantic and all congruously related to the target word being probed. For this latter group, all cues were semantic regardless of the nature of the original encoding question. It is suggested that if in fact subjects have available for all items a minimal semantic representation and further that semantic retrieval information is a relatively more effective retrieval cue, then the latter group of subjects should show an increase in overall retention on the cued recall test. More specifically, this improvement would be expected to reveal itself primarily in the retention of those items coded at lower levels.

Method

Subjects and Design

Sixteen subjects were randomly assigned to each of three independent groups defined strictly in terms of the nature of the retrieval cues presented on the memory test. Group 1 was presented cues which were identical to the original encoding questions. Group 2 was presented cues which were always different from the original encoding questions. However, the cues this group saw were always from the same level of coding as the original encoding questions and were also congruously related to the probed items. The cues given to Group 3 were always semantically and congruously related to the targets and, as for Group 2, each retrieval cue was different from the original encoding question. Examples of each are shown in Table 6. It will be noted there that incongruous structural items were not probed in Group 2. The reason for this was purely practical: it was felt that giving subjects a new structural cue might lead them to generate a target which was other than the one being probed. Thus it would be possible for subjects to generate a target word from some other encoding condition and thereby artificially lower the estimate of retention for this other encoding condition. In retrospect, it might have been equally reasonable to have eliminated structural encodings from the experiment altogether since similar problems could have arisen in Groups 1 and 2 for both congruous and incongruous structural items. Fortunately, however, this was not done since the findings, particularly for Group 3, are very revealing.

Lists

Subjects were presented a total of 54 encoding question-target pairs of which the first six were practice items. These latter items were not included in any of the data analysis.

Each target word was presented in two different encoding conditions across subjects. Further, two different encoding questions were used in each of these conditions resulting in essentially four different lists. An equal number of subjects in each of the three groups saw each list. The reason for using two sets of encoding questions for an item in a given encoding condition was that it permitted half of the subjects in Group 2 to encode with one set and to be tested with the other while for the other half of these subjects, the encoding questions and test cues were reversed.

Procedure

All groups were presented with the standard orienting task under incidental learning instructions. After the delay interval had been completed, all subjects were read cued recall instructions which explicitly identified the nature of the test cues as they related to the previous encoding questions. A number of examples were given directly from the list although these items of course were not subsequently tested. The experimenter attempted to ensure that each subject understood the test cue-target relationships. Of the eight experimental items presented to subjects only six were subsequently tested: the first and last items presented in each of the six encoding conditions were not tested. Subjects were given a minimum of 5 min to complete the recall test although they were permitted to take longer if they had not finished at that point. The test instructions again advised subjects to establish a low output criterion for any targets about which they were not sure.

Results

Reaction times. Analysis of RT's was performed on only the six critical items which were tested in the cued recall phase of the experiment. The results showed main effects due to level of coding ($F(2, 1575) = 54.03$, $p < .01$) and replications ($F(5, 1575) = 9.10$, $p < .01$). Structural questions were answered faster than either phonemic or semantic questions ($M = 0.847$, 0.972, and 0.990 sec respectively). The replications effect is again a practice effect. All other main effects were non-significant. The only significant interaction was between test condition and response type ($F(2, 1575) = 9.78$, $p < .01$). This is attributable to the third group requiring longer to answer congruous than incongruous questions, in contrast to the typical effect which was observed in the other two groups. Since test condition was only revealed subsequent to the orienting task, it must be presumed that this is a random effect. Parenthetically, it was noted that all but one of the interactions involving replications yielded _F_'s of less than 1.0, indicating that item effects had perhaps been reduced to a minimum compared to the previous experiments.

Cued Recall. An incomplete factorial design resulted from the fact that incongruous structural items were not probed in Group 2. Thus an analysis of variance was performed on only those five remaining encoding conditions which were tested in all groups. Descriptively, subjects in Groups 1 and 3 recalled only 2.78% and 6.48% of all incongruous structural targets probed on the test trial with no subject in either condition contributing more than one target.

The analysis of the other five encoding conditions revealed no main effect due tc test condition ($F < 1$, $MSe = 1.989$). There was a main effect of encoding condition ($F(4, 180) = 161.18$, $p < .01$) verifying the large retention differences between congruous semantic items at one end and congruous structural items at the other end. This effect is illustrated in Figure 3 where it is evident that the typical levels x response type interaction is the basis for this main effect. Finally, the interaction of encoding and test condition was also significant ($F(8, 180) = 3.52$, $p < .01$). This effect is attributable to Group 1 performing better on congruous phonemic and congruous semantic items but worse on incongruous phonemic items than Groups 2 and 3. This interpretation was supported by the results of a subsequent analysis involving just the latter two groups. The interaction component was non-significant ($F < 1$, MSe = 0.856) in this analysis. The error term for both of the significant effects reported above was 0.854.

Discussion

Experiment 3 was designed to assess the legitimacy of the concept of a minimal core representation. Given that semantic cues are more effective for retrieval and that some semantic information was in fact available in the memory trace, it was hypothesized that Group 3 would recall more targets than the other two groups. This effect would be restricted to lower level items since it is for these items that the latter groups were using relatively less efficient retrieval cues compared to Group 3. Group 2 was included in this experiment for the purpose of identifying encoding specificity effects due to the use of different input and output cues. Thus, while subjects were fully informed as to the nature of the new cues as they related to the targets being probed (cf. Santa & Lamwers, 1974), it was felt that Group 2 might provide the most appropriate comparison for the interpretation of recall of Group 3 since both groups were presented new output cues at test.

The results showed clear encoding specificity effects for congruous phonemic and congruous semantic items. Group 1 recalled more targets in both encoding conditions than did Group 2 although the relationship between the targets and the retrieval cues was identical for both groups. A further comparison of these two groups revealed large retention differences for incongruous phonemic items: Group 1 recalled only 0.93% of all targets in this condition while Group 2 output 16.67% of these items. In one sense, this is an encoding specificity effect in reverse in that the new cue is more effective than the old cue. Obviously this is not unexpected on a logical basis since Group 1 used an incongruous cue on the test trial whereas Group 2 used a congruous cue. Performance by these first two groups on congruous structural items was too low to make a meaningful comparison. Failure to find a facilitation in the recall of incongruous semantic items by Group 2 (and Group 3) is curious. It seems that the original incongruous semantic cues used by Group 1 are just as effective for retrieval as new but congruous semantic cues. This finding is difficult to interpret at an intuitive level. One possibility is that subjects were able to generate an idiosyncratic congruous relationship between what were nominally incongruous encoding-target pairs. This might artificially raise the retention levels of Group 1. The only evidence against this suggestion comes from the RT data. If subjects were performing in this way, we might potentially expect more errors to appear in the RT data for this condition: subjects would respond "yes" rather than "no". However, error rates were no higher in this encoding condition than in any other and thereby argue against this alleged process.

The critical data in terms of the minimal core representation derives from a comparison of Groups 2 and 3.

In both groups, floor effects were apparent in the recall of structural items. While initially this might suggest that no interpretation is possible, it must be noted that Group 3 was cued with semantic retrieval information. Still they could not access the target items. That even semantic cues were ineffective perhaps suggests that the memory traces for these items were very weak indeed and potentially unavailable on the whole.

A comparison of Groups 2 and 3 for retention of phonemically-encoded items is the most interesting of all. For these items, recall levels were sufficiently above zero for group differences to appear. While small, the obtained differences were consistent. Group 2 recalled more congruous phonemic items than did Group 3 (17.59% and 12.96% respectively) and also more incongruous phonemic items (16.67% and 12.96% respectively). Conceptually, the only difference between Groups 2 and 3 was that the former was presented retrieval cues which were from the same level as the original encoding questions (phonemic) while the latter group used cues from a different (semantic) level. It may be, then, that the best retrieval cue is one which derives from the same level as the original encoding. A complete factorial manipulation of encoding and retrieval cues at all levels would be necessary to fully validate this statement.

Theoretically, it is also important to note that while group differences were small, they were not even in the hypothesized direction. It seems that semantic retrieval cues are not uniformly more efficient than lower level cues as was implicated by Craik and Tulving (1975).[1*]

Groups 2 and 3 used exactly the same set of new congruous semantic retrieval cues for the semantically-encoded items. Thus it is not surprising that performance levels for these items did not vary to any great extent. The only difference between the two groups in terms of these items was that, by the nature of the cue, Group 2 knew (or at least could have figured out) that they were searching for a target which had previously been encoded semantically. Group 3 could not derive this information since all cues were semantic. Apparently this did not affect recall of the targets. Finally, we might mention again, in passing, the surprising failure to find any improvement in retention of incongruous semantic items by Groups 2 and 3 (using new congruous cues) compared to Group 1 (using the previous incongruous cues).

Experiment 4

At this point, the evidence for a minimal core encoding at the semantic level is not overwhelming. In this next experiment, we will attempt to provide a better opportunity for the semantic trace to reveal itself. By using the somewhat more sensitive recognition testing procedure, it was hoped that the existence of a semantic memory component would become evident for items encoded with a lower level orienting task.

[1*] In the earlier quotation from Craik and Tulving, it may be noted that reference was made only to semantic analyses being more beneficial for recall. However it follows that semantic retrieval cues must also be more beneficial (given that a semantic representation exists) since in most circumstances a semantic retrieval cue can access only semantic information. The (probably) rare empirical exception to this statement would be the case in which the subject recalls a target with some retrieval information other than that specified by the experimenter and then tries to match the target to one of the "cues" given to him.

The study procedure for the first two groups was identical to the last experiment. Subjects in these groups differed only in terms of whether they received incidental (Group 1) or intentional (Group 2) learning instructions. A third group did not perform the orienting task but was simply instructed to learn the words for a subsequent memory test. Subjects were then presented a 3-AFC recognition test in which the distractors were unrelated, phonemically-, or semantically-related to the target word. This manipulation was factorially combined with type of orienting task (phonemic, semantic). It was predicted that if subjects automatically encode semantically even though presented a phonemic orienting task then performance should decline when the test distractors are phonemically- or semantically-related to the targets. Similar effects should occur when a semantic orienting task is specified since phonemic analysis is apparently a necessary step in arriving at the semantic stage. The decrement in performance should be measured in terms of recognition levels for targets embedded within unrelated distractors. This analysis is based on the results of previous research showing such detrimental effects of high similarity distractors (e.g., Anisfeld &

Knapp, 1968; Underwood & Freund, 1968).

Method

Subjects and Design

Eighteen subjects were assigned to each of the three independent groups which were differentiated on the basis of the nature of the study conditions. Group 1 was given the standard orienting task under incidental learning instructions. Group 2 was identical to Group 1 except that learning was intentional. Group 3 did not perform the orienting task but rather were simply presented the targets and instructed to learn them.

Lists

A total of 44 items were presented to all subjects of which the first and last four were designated as buffer items. These items were not included in any of the analyses. Two lists were used although actual target items were presented in exactly the same order in each list. The lists differed in terms of the encoding question associated with each respective target. This of course was a pseudo-variable for Group 3.

Procedure

The study procedure for Groups 1 and 2 was exactly in accord with that outlined in the General Method. Only phonemic and semantic questions were used in this experiment. Group 3 was presented the words individually at a 3 sec rate. All groups then performed the standard delay task.

After the delay interval, subjects were read the test instructions which described the 3-AFC recognition procedure in detail. These instructions also indicated that subjects were to indicate the degree of confidence which they held in their recognition decision. A 1 - 4 scale was used as in the earlier experiment involving confidence judgments. Distractor type was not mixed as in previous research (e.g., Underwood & Freund, 1968). Rather, all distractors paired with a given target word maintained exactly the same type of relationship with the target. Thus distractors were all phonemically-related (DUCK-LUCK-TRUCK) semantically-related (CAR-TRAIN-TRUCK), or were both phonemically- and semantically-unrelated to the target (NOUN-STAR-TRUCK). Two sets of recognition lists were constructed using all 36 experimental items. The test lists varied according to the type of distractors associated with each target. Subjects were given as much time as necessary to complete the test.

Results

Reaction times. Since Group 3 did not perform the orienting task, RT data were only available from Groups 1 and 2. The analysis revealed significant main effects due to response type ($F(1, 1190) = 10.49$, $p < .01$) and replications ($F(8, 1190) = 4.18$, $p < .01$). These effects reflect the typical observation that "yes" responses were somewhat faster than "no" responses (M = 1.06 and 1.11 sec respectively) and the effects of practice across trials. The pooled within-subject error term was 0.072. There was no main effect of levels in this analysis ($F(1, 1190) = 2.25$, $p > .10$). In retrospect this should not be entirely unexpected since in the previous studies the levels effect was primarily a function of structural questions yielding much faster RT's than either phonemic or semantic questions which themselves did not differ greatly. In this experiment, only phonemic and semantic questions were presented and again only small differences were observed ($M = 1.07$ and 1.10 sec respectively). There was also no main effect due to study instructions ($F(1, 34) = 2.32$, $MSe = 2.481$, $p > .10$). The only interactions observed were attributable to item effects associated with the replications factor. All other effects were not significant (all p's $> .10$).

Recognition. The first analysis involved recognition performance for all three groups using encoding condition as a pseudovariable for Group 3. This analysis yielded a main effect of study condition ($F(2, 51) = 11.61$, $MSe = 1.084$, $p < .01$) supporting the observation that Group 3 correctly recognized somewhat more targets (77.64%) than either Groups 1 (62.08%) or 2 (67.36%). A Duncan's multiple range test revealed no significant differences among any of the groups (all p's $> .05$).

There were also significant main effects of levels ($F(1, 561) = 42.61$, $p < .01$) and response type ($F(1, 561) =$

28.53, $p < .01$). The error term for these and all subsequent effects was 0.530. Semantically-encoded items were recognized better than phonemically-encoded items (74.17% and 56.81% respectively) and congruous items were recognized better than incongruous items (72.36% and 58.61% respectively). There was, however, no main effect due to type of distractor ($F < 1$). The nature of the distractors used on the test trial did not influence recognition performance.

Only two interactions attained significance. These were the effects of study condition x levels and study condition x response type ($F(2, 561) = 13.56$ and 8.19 respectively, p's < .01). These effects are largely due to levels and response type being pseudovariables for Group 3 and thus the typical depth and congruity effects did not materialize for this group in contrast to the other two groups. The effect of levels x response type approached but did not attain significance ($F(1, 561) = 3.17$, $.05 < p < .10$). No other interactions approached significance (all p's > .10).

A second analysis was carried out on these data in order to confirm which of the above effects were indeed due to the fact that Group 3 did not perform the orienting task. This additional analysis involved only Groups 1 and 2. The main effect of study condition was diminished in this analysis ($F(1, 34) = 1.71$, $MSe = 1.060$, $p > .10$). All subsequent effects were based on a pooled error term of

0.637. The main effects of levels and response type were both magnified ($F(1, 374) = 55.89$ and 34.91, $p < .01$). This would be expected since Group 3 tended to reduce both of these effects in the earlier analysis. The nature of the distractors used in the recognition test again had no effect on performance ($F < 1$).

The previously observed interactions of study condition x levels and study condition x response type did not approach significance in the present analysis ($F(1, 374) = 2.09$ and 2.46, $p > .10$). This finding supports the earlier suggestion that those effects were attributable to the pseudoclassification of levels and response type in Group 3. As with the main effects of levels and response type, the statistical interaction of these two variables was increased when Group 3 was not included in the analysis ($F(1, 374) = 5.25$, $p < .05$). This interaction is quite typical of test performance following the orienting task procedure. All other interactions were non-significant (all F's < 1).

Confidence Judgments. No data analysis was performed on confidence judgments due to the small number of observations contributed by each subject to the various cells of the data matrix. This led to many empty cells and thus a d' analysis was not possible. However, the findings will be reported here in descriptive form. The data are summarized in Table 7. It seems clear from observation of these data that confidence judgments simply mirror each of the effects found in the analysis of recognition performance. For theoretical purposes, it is important to note that type of distractor again had no consistent effect on the data.

In the earlier analysis, we observed a small facilitation in the performance of Group 2 over Group 1, the two groups differing only in terms of the learning instructions. This slight facilitation was also apparent in the confidence judgments for items correctly recognized. Curiously, though, this effect of groups was also evident in the confidence judgments assigned to the false alarms. It is possible, then, that the instruction to learn may have had an overall effect of raising the belief subjects held in their performance capabilities. The alternative view that subjects given incidental learning instructions may have had little faith in their ability to perform the memory task is at least equally tenable based on casual observation of our subjects' reactions to the announcement of the test trial. These possibilities might be considered in future work using incidental and intentional learning instructions.

Discussion

It was hypothesized that subjects automatically analyze highly familiar material to a relatively deep semantic level and as a result this semantic information becomes a part of the memory trace for the event. This automatic processing of semantic information occurs despite the orienting task requiring lower level (e.g., phonemic) information. If this semantic information is in fact available, we would expect to see some decrement in recognition performance when each distractor item shares common semantic elements with the target item. If semantic information is not available following a phonemic orienting task, then phonemic similarity among distractors should result in a performance decrement but semantic similarity should have no effect compared to a set of unrelated distractors.

That neither analysis revealed even the slightest hint of a performance decrement due to the nature of the similarity among distractors might at first suggest that the hypothesis was confirmed in favour of a core representation. Thus performance in the presence of phonemic and semantic distractors was not differentially affected. Unfortunately, however, such a conclusion must be treated with considerable caution since this failure to find an effect due to type of distractor also means that performance with high similarity distractors was not reduced relative to the control condition in which distractors were unrelated to the target. This was an unexpected finding which limits the interpretation of the data.

It has recently been suggested to the author[15] that the nature of the recognition task may be the basis for this empirical failure: in some circumstances, apparently, the m-alternative forced choice (m-AFC) procedure does not result in the predicted effects due to similarity of the distractors. Rather, the effect may only materialize under the yes/no procedure. Of the two studies cited earlier as the basis for the empirical prediction, it was true that Anisfeld and Knapp (1968) used the yes/no procedure. However, the Underwood and Freund (1968) design involved the m-AFC technique and they too found that subjects made more false recognition responses to semantically-related words than to the unrelated control items. It should be pointed out, though, that in this earlier study type of distractor was mixed within any one set of items from which a target was to be selected. Thus, a target was paired with three distractors, one of which was semantically-related, another was formally similar, and the final one was unrelated to the target. In the present experiment, all distractors within any one given set of items were either phonemically- or semantically-related to the target. It seems possible that the task used by Underwood and Freund might have predisposed the finding of a difference in false alarms since only if subjects knew absolutely nothing and merely guessed on a random basis would differences have not occurred.[16]

---

[15] I would like to thank R. Fisher and M. Humphreys for pointing this out to me.

The yes/no procedure provides a rather different situation for the subject. Since many related items will normally precede the targets with which they are associated, they might well be identified as targets if subjects are operating on the basis of partial information and/or they establish a low criterion for acceptance. In contrast to the m-AFC procedure, the yes/no task would seem to maximize the potential for these types of factors to operate. It is perhaps the case, then, that if subjects are required to attend more completely to the recognition items, performance decrements due to similarity among distractors might not be evidenced. The m-AFC procedure could be an example of such a situation.

---

[16] Very recently, Coltheart (1977, Experiment 2) has reported no differences in error rates for phonemically- and semantically-related distractors as a function of encoding condition. These distractors were mixed, as in the Underwood and Freund study, although in most other respects the experiment was very similar to the present one.

In all, the present experiment may not have provided a particularly good test of the concept of a core representation as was suggested earlier. More complete parametric investigations of the conditions under which distractor items lead to performance decrements seem warranted. Until these conditions have been elaborated, the findings of the present study must be held in abeyance.

Experiment 5

The final experiment was designed with somewhat more parametric purposes in mind than each of the earlier studies. The question asked was very simply whether or not subjects could learn to overcome the typical depth effect if they were given multiple trials using exactly the same orienting task procedures. In other words, would retention for items encoded at lower levels in the system show an improvement after some practice in trying to retrieve these items? Further, if retention did improve, could it be attributed to subjects actually learning to retrieve these lower level encodings or would it be a function of additional deeper level processing of these items at input? The latter possibility seems entirely reasonable since subjects would become very aware of the fact that their recall protocols in the first trial or two were composed almost entirely of semantically-encoded items: few lower level items were being retrieved. Thus, since subjects were given plenty of time to process items any way they wished subsequent to performing the orienting task, it seems likely that they might choose to develop a somewhat more elaborate semantic code for these items. In this way, they would have available to them more substantial and efficient retrieval information which would facilitate recall relative to the lower level information derived from the orienting tasks.

Two groups of subjects were presented the standard orienting task on each of three trials. Learning was intentional on all trials. One group was given free recall on each trial while the other group was given a cued recall test. It was expected that the cued recall group would have no reason to elaboratively encode lower level items in the semantic domain since they were presented the original cues on the test trial: nothing is to be gained in having a strong semantic trace if the retrieval cue specifies structural or phonemic information. On the other hand, the free recall group, who generate their own retrieval cues, would benefit maximally by such extended deeper level processing since they could use this on the test trial when they attempt to reconstruct the target items.

It is possible, though, that practice in using the lower level encoding questions might serve to increase their cueing potential. This would result in the cued recall group also showing a performance increment across trials in this experiment. There is no theoretical basis for making this prediction based on any version of depth of processing. Rather, a rigid theorist would seem to be relegated to the belief that lower level cues are invariable ineffectual at producing recall, regardless of the amount of practice allotted to them.

Now it is of course true that both the free recall and cued recall groups could show a performance increment across trials on the latter basis. Thus a simple comparison of these two groups might well lead to an ambiguous conclusion since we would not know whether the free recall subjects had processed items deeper and subsequently used that information at retrieval or whether they, like the cued recall subjects, simply benefited from practice at using the lower level information. It was necessary then to have some independent measure of assessing whether the free recall subjects were in actuality processing to relatively deeper levels. To this end, two additional groups were run. Each was given just the third of the three lists presented to the other two groups.

After all groups had performed the free recall test on each of one or three trials, they were given a final cued recall test. In most ways, this test was similar to that given to subjects in Experiment 3 reported earlier. Subjects were presented a series of semantic cues which were all new and all congruously-related to the target word being probed. If the free recall subjects were learning over trials to develop a more elaborate semantic memory component, then we might expect them to perform somewhat better on this final cued recall test in comparison to the free recall group given but a single trial. This should be particularly true of items encoded with the lower level orienting questions.

Method

Subjects and Design

Twenty subjects were randomly assigned to each of four independent groups. Groups 1 and 2 (1-trial groups) were given a single trial using the orienting task procedure and were tested under free recall and cued recall conditions respectively. Group 3 and 4 (3-trial groups) were presented three trials and were tested on each list with, again, free and cued recall respectively. All groups were given intentional learning instructions but were not informed of the final cued recall test until it was actually given to them.

Lists

A total of three lists were constructed, each composed of 36 question-target pairs. The orienting task consisted of structural, phonemic, and semantic questions. The first list presented to the 3-trial groups commenced with an additional six practice items as did the only list presented to the 1-trial groups. These items were not included in any of the data analyses. The 1-trial groups were always presented with the third list which the 3-trial groups were given. This permitted a direct comparison of performance on the final cued recall test since all subjects were being tested on exactly the same set of words. As in the previous experiments, all targets were presented in exactly the same order although the actual encodings associated with each target varied for different subsets of subjects. Two different sets of encodings were used on each list.

Procedure

All groups were fully informed prior to the beginning of the experiment how many lists they would be seeing and that they would be tested on each list shortly after the last item had been presented in each list. Since the 3-trial groups required three delay intervals, one after each list, it was necessary to institute a different procedure. All groups were given a page of simple arithmetic problems to complete during the 90 sec delay interval. It was stressed that their performance on this test would be evaluated and that they should therefore use the 90 sec to complete as many problems as possible without sacrificing accuracy. This instruction was included in the initial set of instructions at the outset of the experiment and again when subjects were actually handed the problems after each study trial. It was hoped that this would minimize rehearsal of targets during the delay interval.

For the test trial, free recall subjects were simply given a sheet of paper containing only blank lines. The cued recall subjects were given exactly half (3 of 6) of the original questions from each of the six encoding conditions and were asked to recall the associated target items. These tasks were fully explained at the outset of the experiment. In addition, the experimenter gave a brief informal description of the subject's task on the first test if it was necessary. All groups were given a minimun of 3 min to recall as many targets as possible.

After subjects had completed the test trial for the last list, they were engaged in the standard delay task of completing the relevant administrative cards associated with their experimental participation. Subsequent to this, they were read instructions for the final cued recall test. By the use of a number of examples, it was indicated in these instructions that the cues in this test would all be either phonemic or semantic and would all be congruously related to the target. Rather than attempting to specify all the possible relationships between the original encoding conditions and the final test cues, the instructions simply tried to convey to subjects that all combinations were possible. This was not in fact the case since semantically-encoded items were never probed with phonemic cues: there did not appear to be any logical theoretical basis for including these types of relationships in the test. Further, test cues were identical to the original encoding questions in those situations in which both were congruous phonemic or congruous semantic: different cues were not used in these situations (to control for encoding specificity effects) since they were not of particular interest. Rather, the focus of the final cued recall test was on items for which the test cues reflected a deeper level of analysis than that required by the original encoding question. Two items were probed in each of the 10 encoding condition x test condition combinations. Subjects were given as much time as they required to complete this test.

Results

Reaction times. The RT data for the 1-trial groups will be presented first followed by that of the 3-trial groups. The main effect of test condition was nonsignificant ($F < 1$, $MS_e = 2.710$). Each of the other main effects, however, was significant. These and all other effects were tested with a pooled error term of 0.082. The main effects of levels ($F(2, 1330) = 37.23$, $p < .01$), response type ($F(1, 1330) = 18.42$, $p < .01$), and replications ($F(5, 1330) = 5.16$, $p < .01$) were entirely typical and need not be elaborated. The only interaction of note was a small effect of levels x response type ($F(2, 1330) = 4.39$, $p < .05$). As in Experiment 2, this was entirely due to the phonemic questions: "yes" and "no" responses yielded equal RT's for phonemic questions whereas RT's for "yes" responses were on the whole larger than for "no" responses in the structural and semantic encoding conditions. No other interactions were observed which could not be attributed to item effects associated with the replications factor.

Analysis of RT's for the 3-trial groups was complicated by the fact that two of the error components deviated significantly from the rest. Thus, it was not theoretically justifiable to pool all of the within-subject error terms. The appropriate error component will be presented along with the report of each statistical effect.

There was no main effect on RT's of test condition ($F(1, 38) = 1.63$, $MSe = 11.123$, $p > .10$). The main effect of trials was also non-significant ($F < 1$, $MSe = 0.311$) indicating that any practice effects were overcome early in the first trial. Replications yielded a small effect ($F(5, 190) = 2.81$, $MSe = 0.094$, $p < .05$) however this is more an effect of individual items than it is a practice effect with trials. The main effects of levels ($F(2, 76) = 69.51$, $MSe = 0.089$, $p < .01$) and response type ($F(1, 38) = 14.32$, $MSe = 0.033$, $p < .01$) were also clearly evident.

While a few interactions were obtained, all but two appeared to be due to item effects. The interaction of levels x response type ($F(2, 76) = 6.94$, $MSe = 0.095$, $p < .01$) is again a function of differential performance with the phonemic questions: relatively large differences in RT's for "yes" and "no" responses to structural and semantic questions were substantially reduced in the context of phonemic questions. The second interaction of interest was trials x levels (_F_(4, 152) = 4.32, _MS_e = 0.076, _p_ < .01). It appeared that RT's to semantic questions decreased across trials whereas RT's to structural questions showed a slight but inconsistent increase. There was no change in RT's to phonemic questions.

Immediate Recall Performance. The first hypothesis of interest was whether the free and cued recall groups would show an increase in immediate recall across successive trials. The data were analyzed in terms of the actual recall scores and thus it must be noted that the cued recall groups were only tested on half of the targets whereas the free recall groups were instructed to attempt recall of all targets.

The results are shown in Table 8 for the 3-trial groups. The main effect of test condition was significant (_F_(1, 38) = 12.70, _MS_e = 3.426, _p_ < .01) with the free recall subjects recalling more targets than the cued recall subjects. In terms of percent recall, the cued recall group output 28.61% of the targets probed whereas the free recall group retrieved 22.50%. All other statistical effects were based on a pooled error term of 0.606. There was a main effect of trials (_F_(2, 646) = 7.26, _p_ < .01). Performance increased by more than 26% from trial 1 to trial 2 but then decreased by about 4% from trial 2 to trial 3. There is a suggestion then that performance might asymptote as early as the second trial with the effects of experience with this procedure limited to the first trial. Based on the substantial amount of previous research, the main effects of levels ($F(2, 646) = 199.87$, $p < .01$) and response type ($F(1, 646) = 156.05$, $p < .01$) were very much expected.

Two hypotheses were advanced earlier with regard to performance changes across trials. One suggested that the free recall groups might choose to develop more effective semantic information on later trials and the other involved an increase in the cueing potential of the lower level encodings. If the former effect was operating but not the latter, we would predict an interaction between test condition and trials since only the free recall groups could capitalize on additional deeper level processing. If cues become more effective with practice, we might predict no interaction because subjects in both test conditions could make use of this. A similar prediction would be derived if the first factor was operating in the free recall group and the second factor was in effect for the cued recall group. Indeed, the data showed no interaction of test condition x levels ($F = 1$). The final cued recall test was designed to help ascertain which of the latter two hypotheses was most tenable. This will be reported later.

The first order interaction of levels x response type revealed a highly significant effect ($F(2, 646) = 55.85$, $p < .01$). Congruous items were retrieved much better than incongruous items after a phonemic or semantic orienting task but there was no difference at the structural level. This effect was revealed in both the free and cued recall data as was indicated by the failure to find a significant second-order interaction of the three variables ($F(2, 646) = 1.39$, $p > .10$). Two other interactions were observed, test condition x levels ($F(2, 646) = 3.36$, $p < .05$) and test condition x response type ($F(1, 646) = 6.43$, $p < .05$). The advantage of the free recall group in terms of absolute recall decreases with deeper level encodings. Also, the free recall group retrieved disproportionately more incongruous items than the cued recall group. No other effects approached significance (all p's $> .10$).

The analysis of immediate recall performance by the 1-trial groups is of little theoretical interest in the present experiment and thus will be considered only briefly. There was no main effect of test condition ($F(1, 38) = 2.92$, $MSe = 0.892$, $p > .10$) although, in terms of percent recall, the free recall subjects output 15.56% and the cued recall subjects 24.17% of the targets. Main effects were observed for levels ($F(2, 190) = 89.41$, $p < .01$) and response type ($F(1, 190) = 37.14$, $p < .01$) reflecting the entirely typical effects of depth and congruity. These and subsequent effects were tested with a pooled error term of 0.474.

The first order interaction of test condition x response type was significant ($F(1, 190) = 17.80$, $p < .01$). As with the 3-trial groups, this effect is due to the relatively few incongruous items retrieved by the cued recall group compared to the free recall group. Disproportionately more items were recalled from the congruous semantic encoding condition resulting in an interaction of levels x response type (F(2, 190) = 23.58, $p < .01$). Finally, a small second-order interaction of test condition x levels x response type was observed (F(2, 190) =3.40, $p < .05$). Comparatively few incongruous items were retrieved by the cued recall group (particularly from the lower level encoding conditions) while concurrently this same group was recalling more congruous semantic items. Essentially the same interpretation would describe these data were they presented in terms of percent recall.

Final Cued Recall. The final data analysis involved a comparison of all groups on the final cued recall test. It was hoped that this test would permit a separation of the two theoretical interpretations offerred earlier to account for the performance increments across trials noted for both the free and cued recall groups.

The data from this recall test are presented in Figures 4 and 5 for the phonemic and semantic cues respectively. Separate analyses were performed on the data resulting from these two types of cues. Analysis of retention in the presence of the phonemic cues revealed a small effect due to the testing condition used in the immediate recall trial (F (1, 76) = 6.09, MSe = 0.432, $p < .05$).[17] Subjects who had previously had free recall tests retrieved more targets on this final test than subjects previously exposed to cued recall (23.75% vs 14.69%). The effect of the number of trials on which subjects had been tested approached significance ($F(1, 76) = 3.19$, $MSe = 0.432$, $p < .10$). The 3-trial groups tended to recall more targets than the 1-trial groups (22.50% vs 15.94%). The interaction of these two between-subject factors was non-significant ($F < 1$).

All within-subject effects were tested with a pooled error term of 0.312. Phonemically-encoded items were retrieved better than structurally-encoded items ($F(1, 228) = 8.41$, $p < .01$). There was also a slight but non-significant tendency for congruous items to be recalled more often than incongruous items ($F(1, 228) = 2.89$, $p < .10$).

The only interaction component large enough to attain significance was levels x response type ($F(1, 228) = 12.25$, $p < .01$). Retention of congruous phonemic items was much superior to items in all other encoding conditions when the cues were phonemic. Of course, this would not be unexpected since for these items, the encoding and test cues were identical in contrast to items probed from the other encoding conditions. One other interaction approached significance, that of trials x levels x response type ($F(1, 228) = 2.89$, $p < .10$). The 3-trial groups retrieved more targets from all encoding conditions than the 1-trial groups except in the case of incongruous phonemic items. There is

---

[17] The error mean squares used in all of these analyses are based on the raw retention data whereas the data presented in Figures 4 and 5 are in terms of percent recall.

no obvious explanation of this small interaction.

Analysis of the retention data in the presence of semantic cues revealed many similar findings. Main effects were observed for both between-subject variables. The free recall groups recalled 34.38% of the targets compared to 24.17% for the cued recall groups (F(1, 76) = 10.53, p < .01). The 3-trial groups were able to generate 33.96% of the targets compared to 24.58% for the 1-trial groups (F(1, 76) = 8.88, p < .01). The interaction of test condition and trials was not significant (F < 1). The error term for each of these analyses was 0.475.

The pooled within-subject error term used in all subsequent analyses was 0.299. The typical effects of depth of coding (F(2, 380) = 122.31, p < .01) and congruity (F(1, 380) = 27.70, p < .01) were clearly evident. Levels and response type also interacted in the standard fashion (F(2, 380) = 22.86, p < .01).

The analysis revealed an interaction of test condition x response type (F(1, 380) = 9.55, p < .01). In addition, the interaction of test condition x trials x levels approached significance (F(2, 380) = 2.54, p < .10). Both of these effects may be considered in terms of the higher order interaction of all four components, test condition x trials x levels x response type (F(2, 380) = 3.88, p < .05). This interaction is critical in terms of the two theoretical interpretations suggested earlier. As may be seen from Figure 5, the 3-trial free recall group showed a differential improvement in recall with respect to the 1-trial free recall group compared to the cued recall groups. This improvement was restricted to the lower level items. Also involved in this complex interaction were isolated effects associated with the recall of the 3-trial cued recall group. On the whole, the interpretation of this interaction seems to support the theoretical view that the free recall groups were learning to process lower level items for more elaborate semantic information. With this additional information available in memory, these subjects were much more successful at using the semantic cues presented on the final cued recall test.

It might be suggested that the same interaction should have been observed in the presence of phonemic cues, although it very clearly was not ($\underline{F} < 1$). In response to this, though, it should be noted that the phonemically-encoded items were not well retained by the free recall groups under any circumstances. It is unlikely then that subjects would abandon one ineffective lower level code (structural) for another which is only minimally more effective (phonemic). Thus, the failure to observe the interaction in the analysis of cued recall with phonemic retrieval information is not especially surprising.

Discussion

The final study was designed to determine whether subjects could learn with practice to improve their retention of items encoded under the orienting task procedure. Specifically of interest were items encoded with lower level orienting tasks since it is these which are most prone to forgetting.

The results clearly showed that subjects did increase their recall performance with practice, whether tested under free or cued recall conditions. The failure to observe a statistical interaction of testing condition and trials further indicated that the degree of learning was comparable for both groups. This left open the question of exactly what factors were contributing to this effect. The only obvious interpretation of the improvement shown by the cued recall group seems to be that subjects were able to learn how to make more efficient use of the encoding cues. For these subjects to carry out more processing on the targets and thereby develop a more elaborate memory trace would not seem particularly beneficial since the retrieval information used on the test trial is exactly the same as the encoding information. The design of the present study does not permit us to isolate the source of the practice effect associated with the increased efficiency of the retrieval cues. The effect could lie in the encoding/storage stage or it may involve retrieval itself. In any event, it is of interest to note that items from all encoding conditions appeared to benefit from the practice effects although the performance change across trials was not always monotonic.

This interpretation of the recall performance by the cued recall subjects implies an effect in what we have identified as Stage 2 of the reconstruction process. At this point in retrieval, subjects have available the encoding context (retrieval information) and need only activate the target word. The free recall group also showed an increase in recall with practice. Theoretically, it was possible that they too could have been learning to make more efficient use of the retrieval information, once it had been activated. Alternatively, they could have been elaboratively encoding items at a deeper semantic level and thus availing themselves of more efficient retrieval cues for the recall test.

To empirically distinguish between these alternative interpretations, the final cued recall test was included. The results of this test were reasonably clear. The performance of the cued recall group in the presence of semantic cues increased somewhat as a function of trials. The free recall group, however, showed a larger increment, particularly for the lower level encoding conditions. The implications of these findings are that the cued recall group was learning to make better use of the cues given to them whereas the free recall group was learning to elaboratively process items to deeper levels and then use this more efficient information on the retention test.

That subjects can improve their retention of items encoded at all levels does not argue against the depth of processing point of view. Indeed the evidence suggesting that semantic processing is playing an increasing role in the performance of the free recall groups may be indicating that the theoretical relationship between depth and memorability is something which subjects are, or can be made aware of very readily. Further, depth did not interact with trials in the immediate retention of the cued recall groups, suggesting that, at least in the short term, the depth effect cannot be overcome with practice: lower level cues remain less effective than deeper semantic cues.

General Discussion

Each experiment has been discussed extensively and therefore we need not repeat the many details here. Rather, we will summarize the important findings and indicate briefly how they contribute to depth of processing and memory theory in general.

We will deal first with the concept of a core encoding and core representation. In the general Introduction, the views of a number of authors were described as they related to these matters. In brief, the consensus was that much "preattentive" or automatic processing is carried out prior to and without conscious awareness. In this respect, Craik is in agreement with most other authors. Significantly, however, it was concluded that Craik differs on the critical matter of storage of the products of these analyses. In order to maintain theoretical consistency, it was necessary for Craik to hypothesize storage of these preliminary analyses. This is in contrast to each of the other authors cited who explicitly denied that the products of these preattentive analyses were stored.

It may be recalled that a substantial amount of data was cited earlier suggesting that in fact the majority view was most likely correct. To these data, we can now add the findings of Experiment 3 reported herein. When items encoded at lower levels within the processing system were later probed with deeper semantic retrieval information, there was no strong evidence of a substantial semantic memory component. As with all the other data, these results seem to argue against the view that information processed "preattentively" is stored in any long-term fashion.

Depth of processing is not irreparably refuted by this collection of data. Rather the theory needs only to be developed to the point where it clearly distinguishes automatic processing of familiar information from the conscious processing which occurs post-perceptually in accord with the demands of the task. It is a small step, once this qualitative distinction has been made, to attribute long-term retention characteristics to the products of the latter processing only. This would bring depth of processing in line with the many other theoretical views outlined above and the consistent empirical findings in this regard. We have already identified the control process within depth of processing which allows the subject to carry out a deep and elaborate analysis within any processing domain. This control process correlates perfectly with the attentional mechanisms described by each of the other theorists and which permits them to describe how subjects deploy conscious processing capacity to specific aspects of a complex stimulus array.

Our final comments deal with the specific concept of reconstruction. It is important to emphasize that we have established an arbitrary operational definition of both how and when reconstruction operates. Neither cf these uses of the term can be explicitly attributed to Craik and his colleagues although an attempt has been made herein to document the empirical and theoretical rationale used in defining each of these aspects of reconstruction.

The present research provided some interesting insights into the reconstruction process. Possibly the most relevant finding was that subjects who generated their own encodings (Experiment 2, Group 5) virtually never recalled a target word for which they could not also generate the entire encoding. The significance of this result lies in the fact that this group most nearly reflects the "normal" learning situation in which the subject is only provided the TBR. In this way, the data provides perhaps some of the best evidence currently available that reconstruction is indeed a viable retrieval process.

As noted earlier, a number of other authors have also used the term reconstruction in rather different contexts (e.g., Bartlett, 1932; Herriot, 1974; Russo & Wisher, 1976; Weingartner, Walker, Eich & Murphy, 1976) although they have also not defined exactly what processes they are implicating. One interesting application of the term is to the topic of natural language mediation such as discussed by Montague, Adams, and Kiess (1966) and Prytulak (1971). This latter approach seems clearly to be a form of reconstruction, as we have defined it, in that subjects code items in idiosyncratic ways and then use an external cue to permit the reactivation, or reconstruction, of the encoding and the target.

Bartlett's use of reconstruction probably deviates most significantly from that of the other authors. In his well known studies of how people relate stories which they have been told, Bartlett found that few details were in fact remembered. More interesting though was the observation of what subjects did include in their recounting of the stories. In some situations, only an isolated detail would be remembered and subjects put together a rather different story around this one fact. In other cases, subjects were able to recall the basic theme in the story and simply added details which, in many instances, conformed to cultural stereotypes. This latter form of reconstruction is particularly interesting in terms of subjects using their semantic memories, or cognitive structures, to piece together a plausible story (cf. Restle, 1974, for a very compatible approach). A number of other historically relevant studies have been conducted on how subjects use their cognitive structures to facilitate recall (see Kintsch, 1970, pp. 261-7).

Another more recent paradigm to which reconstruction can be applied is the recall of organized conceptual structures. In one study, Bower, Clark, Lesgold, and Winzenz (1969) presented one group of subjects a hierarchically arranged series of words falling within one semantic classification. A second group was presented the same words but randomly arranged within the hierarchical structure. They observed that the former group learned the "lists" much more quickly than the latter group. It could be suggested that the first group was reconstructing the hierarchical arrangement and using that to facilitate their retrieval of the critical target words. For the second group, the hierarchical structure was essentially meaningless and thus there was likely little benefit in using it at the time of recall.

Other empirical paradigms could be described which could reasonably be used as correlated evidence for the existence of a process such as reconstruction. But the point to be made from this selective empirical review is that reconstruction is a theoretical process which is not restricted to a depth of processing framework. Indeed, it can be argued that reconstruction exists in many other approaches, although it is not necessarily identified as such. For example, Humphreys (1976) has recently distinguished the use of item and relational information at retrieval. The latter very readily fits a view of reconstruction as we have described it. Anderson and Bower (1973) have elaborated a complex theoretical model of memory in which retrieval is defined essentially as the re-activation of relational information between target items. If the tagged relational information is considered simply as the encoding context for a given target, than it becomes apparent that this may be another system in which a process of reconstruction is implicit.

Very simply, then, reconstruction as a retrieval process is not theory-bound but rather can legitimately be investigated as a phenomenon in its own right. The depth of processing paradigm happens to be a particularly interesting and potentially fruitful vehicle for this investigation because the encoding context is rather explicitly defined. This is not necessarily true of other empirical paradigms which might otherwise be of interest in this regard (e.g., Anderson & Bower, 1972; Humphreys, 1976).

Considering reconstruction as a retrieval process from a depth of processing point of view raises other very important issues. First, as was mentioned earlier, is the not insignificant matter of distinguishing items retrieved via reconstruction as opposed to either of the other two retrieval processes. Secondly, the reconstruction process itself must be more clearly defined so that it will be possible to determine whether it can account for all long-term retention, as Lockhart et al. (1976) suggest. Both of these matters necessitate a more clear statement of retrieval within the depth of processing framework.

TABLE 1

Experiment 1: Percent Free Recall as a Function of Instructional Condition, Encoding Condition, and Type of Encoding Information Recalled.

GROUP 1: incidental learning

Encoding Information Retrieved	ENCODING CONDITION					
	1-C*	1-I	2-C	2-I	3-C	3-I
Target Only	1.62	0.00	0.88	4.12	0.88	3.38
Target + Level	0.00	0.00	2.50	0.88	0.00	1.62
Target + Encoding	0.00	0.00	5.00	0.00	34.12	6.62

GROUP 2: intentional learning, targets only

	1-C*	1-I	2-C	2-I	3-C	3-I
Target Only	1.62	3.38	0.00	1.62	1.62	4.12
Target + Level	1.62	0.88	3.38	0.88	1.62	1.62
Target + Encoding	1.62	0.00	2.50	0.88	29.12	3.38

GROUP 3: intentional learning, targets and encodings

	1-C*	1-I	2-C	2-I	3-C	3-I
Target Only	0.00	0.88	0.88	2.50	1.62	2.50
Target + Level	1.62	3.38	5.88	4.12	0.88	0.00
Target + Encoding	3.38	0.00	0.88	0.88	34.12	5.00

*1 - Structural encoding; 2 - Phonemic encoding
3 - Semantic encoding
C - Congruous encoding; I - Incongruous encoding

TABLE 2

Experiment 1: Percent Correct Recognition as a Function of Instructional Condition, Encoding Condition, and Type of Encoding Information Recalled.

GROUP 4: incidental learning

Encoding Information Retrieved	ENCODING CONDITION					
	1-C	1-I	2-C	2-I	3-C	3-I
Target Only	15.00	30.00	21.67	40.00	6.67	36.67
Target + Level	35.00	26.67	41.67	28.33	5.00	23.33
Target + Encoding	0.00	0.00	10.00	0.00	83.33	16.67

GROUP 5: intentional learning, targets only

Target Only	10.00	25.00	26.67	35.00	8.33	45.00
Target + Level	46.67	35.00	36.67	26.67	10.00	26.67
Target + Encoding	1.67	0.00	15.00	0.00	75.00	13.33

GROUP 6: intentional learning, targets and encodings

Target Only	8.33	18.33	18.33	31.67	1.67	50.00
Target + Level	38.33	20.00	43.33	28.33	5.00	25.00
Target + Encoding	1.67	0.00	8.33	0.00	88.33	5.00

TABLE 3

Experiment 1: Mean Confidence Judgments for Hits and False Alarms as a Function of Encoding Condition. Data are Averaged across Instruction Conditions.

Response	ENCODING CONDITION					
	1-C	1-I	2-C	2-I	3-C	3-I
Hits	2.62	2.51	3.04	2.87	3.91	3.53
(n)	94	93	133	114	170	145
False Alarms	1.94	2.03	1.83	1.85	1.90	2.06
(n)	83	90	47	66	10	35

TABLE 4

Experiment 2: Percent Recall as a Function of Encoding and Test Condition (Groups 1 and 2).

	FREE RECALL		CUED RECALL	
Encoding Question	Response Type			
	C	I	C	I
Structural	1.50	2.00	0.83	0.00
Phonemic	7.50	4.00	23.33	0.00
Semantic	30.00	9.00	84.17	10.00

TABLE 5

Experiment 2: Percent Recall as a Function of Study and Test Condition (Congruent Encodings Only).

FREE RECALL

	Study Condition		
Encoding Condition	Standard Orienting Task	Generate Targets	Generate Encodings
Structural	1.50	33.25	12.00
Phonemic	7.50	26.00	14.00
Semantic	30.00	45.50	37.25

CUED RECALL

Structural	0.83	59.17	18.10
Phonemic	23.33	81.67	52.76
Semantic	84.17	97.50	88.34

TABLE 6

Experiment 3: Examples of Test Cues used by the Three Groups of Subjects in Each of Three Encoding Conditions.

Example 1: Congruous Structural

Question: Does it begin with the same letter as DEBT?
Target : DOG

Test cues
Group 1 : Does it begin with the same letter as DEBT?
Group 2 : Does it begin with the same letter as DUST?
Group 3 : Is it a type of pet?

Example 2: Incongruous Phonemic

Question: Does it rhyme with EAST?
Target : KING

Test cues
Group 1 : Does it rhyme with EAST?
Group 2 : Does it rhyme with SING?
Group 3 : Is it a member of royalty?

Example 3: Congruous Semantic

Question: Is it a form of precipitation?
Target : SNOW

Test cues
Group 1 : Is it a form of precipitation?
Group 2 : Is it a weather phenomenon?
Group 3 : Is it a weather phenomenon?

TABLE 7

Experiment 4: Mean Confidence Judgments (number of observations included in brackets).

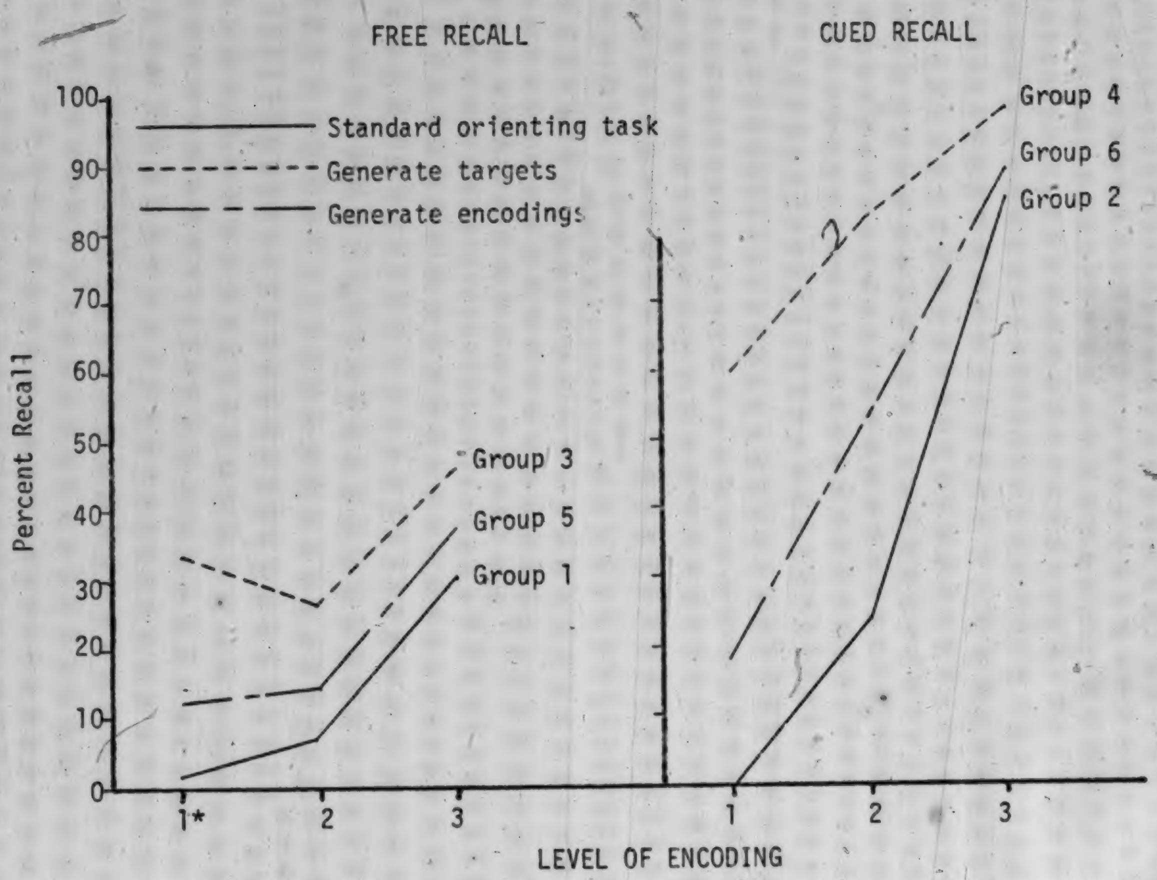
		Group 1			Group 2			Group 3		
		Type of Distractors								
		U*	P	S	U	P	S	U	P	S
2-C**	Hits	2.47	2.41	2.51	2.74	2.84	2.53	3.38	3.44	3.26
		(34)	(39)	(35)	(39)	(39)	(34)	(47)	(45)	(50)
	False Alarms	1.55	1.53	1.58	2.07	2.00	1.80	1.86	1.11	2.25
		(20)	(15)	(19)	(15)	(15)	(20)	(7)	(9)	(4)
2-I	Hits	2.32	1.85	2.30	2.61	2.85	2.48	3.58	3.69	3.57
		(28)	(26)	(30)	(33)	(39)	(33)	(45)	(48)	(46)
	False Alarms	1.50	1.68	1.62	1.71	1.60	1.86	1.89	2.00	2.25
		(26)	(28)	(24)	(21)	(15)	(21)	(9)	(6)	(8)
3-C	Hits	3.73	3.68	3.92	3.71	3.59	3.61	3.64	3.67	3.59
		(52)	(50)	(50)	(49)	(51)	(49)	(47)	(45)	(46)
	False Alarms	1.50	1.50	1.75	2.40	1.67	2.80	1.71	1.78	1.88
		(2)	(4)	(4)	(5)	(3)	(5)	(7)	(9)	(8)
3-I	Hits	3.17	2.91	2.68	2.92	3.24	3.02	3.51	3.50	3.53
		(41)	(35)	(38)	(40)	(38)	(41)	(49)	(46)	(45)
	False Alarms	1.38	1.47	1.69	2.29	1.12	1.69	2.00	2.00	2.78
		(13)	(19)	(16)	(14)	(16)	(13)	(5)	(8)	(9)

*U:Unrelated; P:Phonemically similar; S:Semantically similar
**Encoding Condition: 2:Phonemic; 3:Semantic; C:Congruous; I:Incongruous

TABLE 8

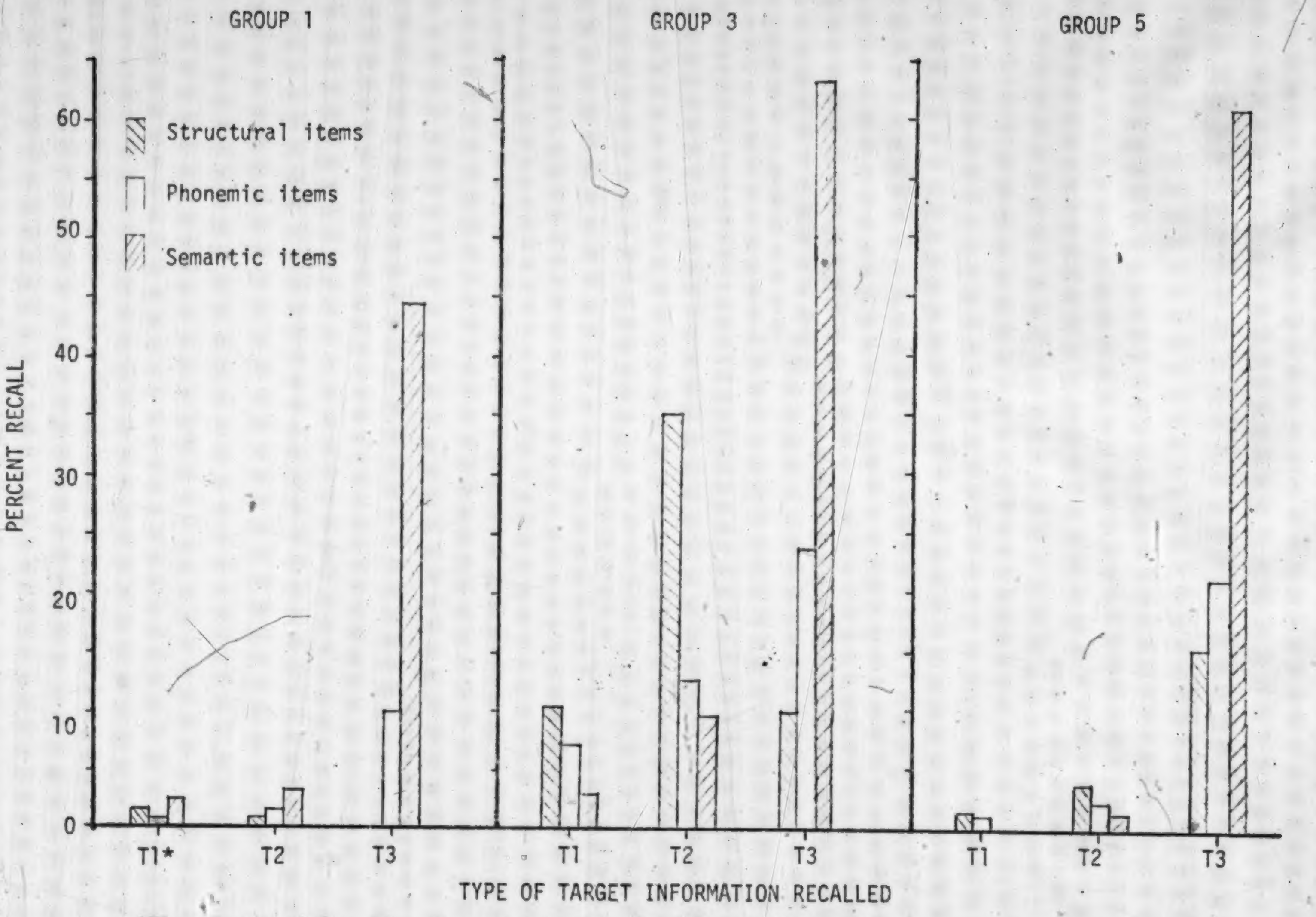
Experiment 5: Percent Recall for the 3-Trial Groups as a Function of Trials and Encoding and Test Conditions.

		Free Recall			Cued Recall		
Encoding Condition		TRIALS					
		1	2	3	1	2	3
Structural	Congruous	10.00	17.50	15.00	3.33	5.00	13.33
	Incongruous	9.17	15.83	20.83	0.00	5.00	10.00
Phonemic	Congruous	21.67	25.00	22.50	35.00	41.67	28.33
	Incongruous	10.00	11.67	18.33	3.33	15.00	10.00
Semantic	Congruous	44.17	45.83	44.17	83.33	88.33	90.00
	Incongruous	20.00	28.33	25.00	25.00	38.33	20.00



*1 = Structural; 2 = Phonemic; 3 = Semantic

Figure 1: Experiment 2: Recall of target words as a function of study condition and level of encoding.



*T1: Target Only; T2: Target & Level; T3: Target & Encoding

Figure 2. Experiment 2: Retention of encoding information by the free recall groups as a function of study condition, level of encoding, and type of target information recalled.

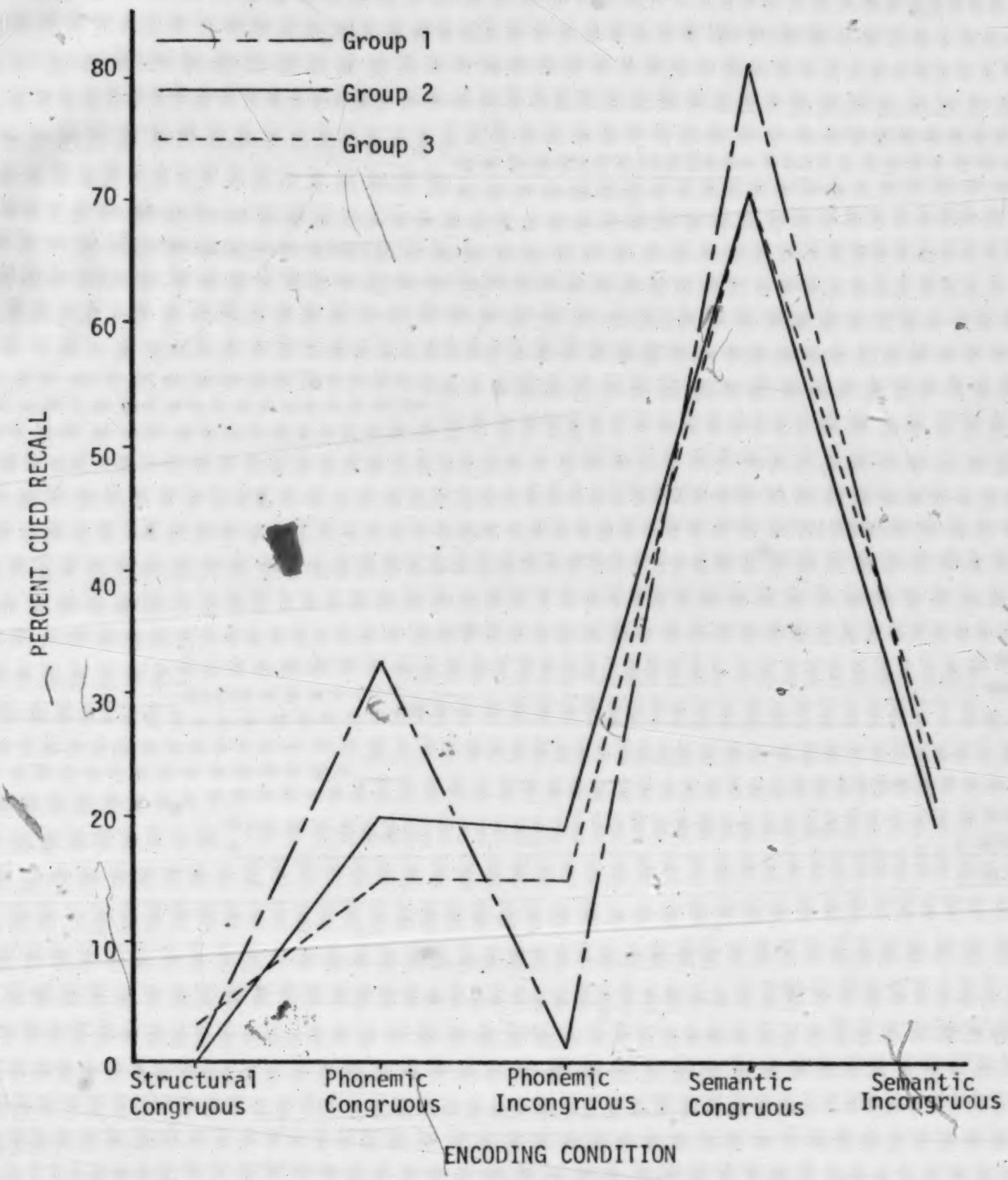
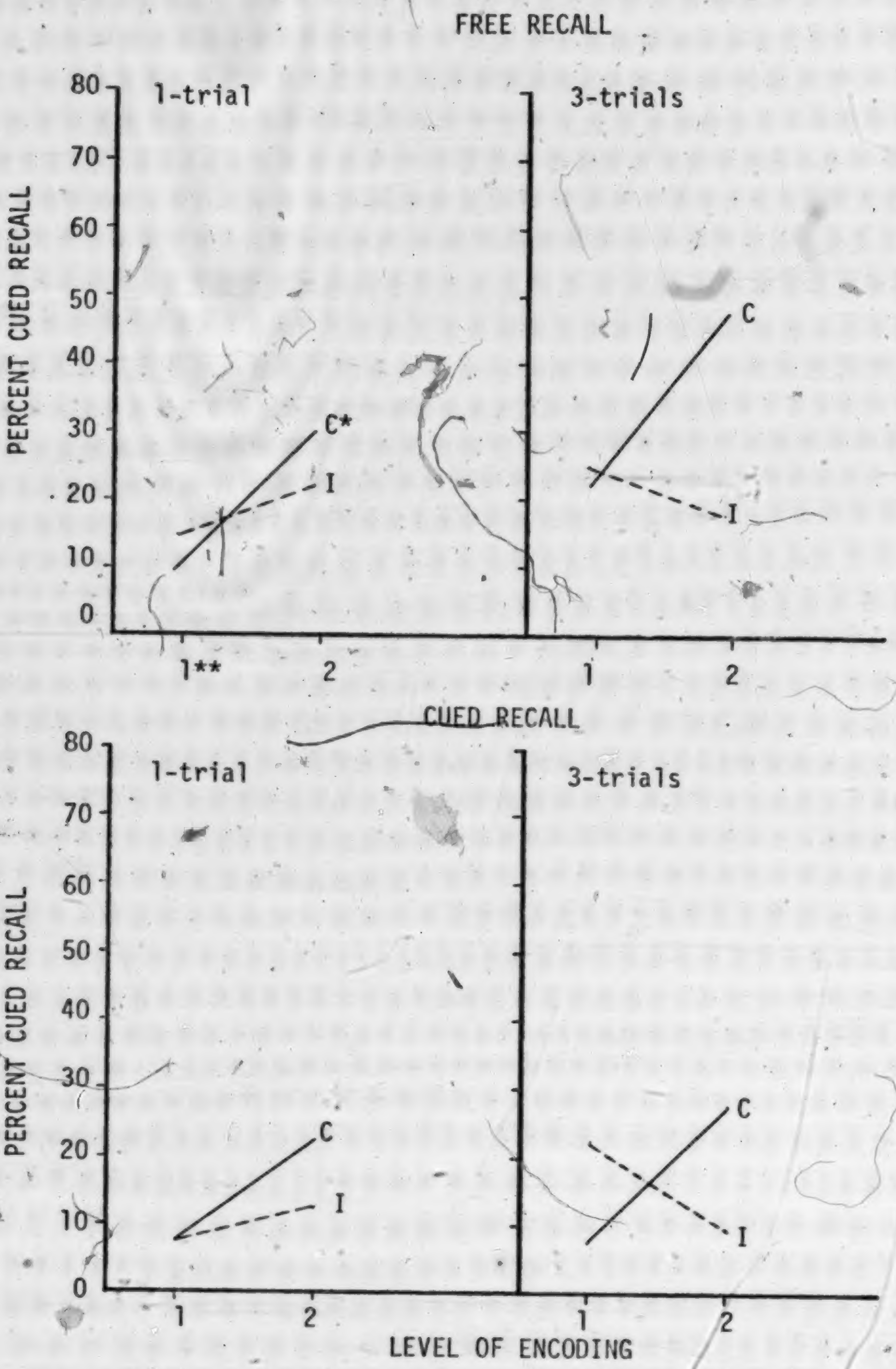


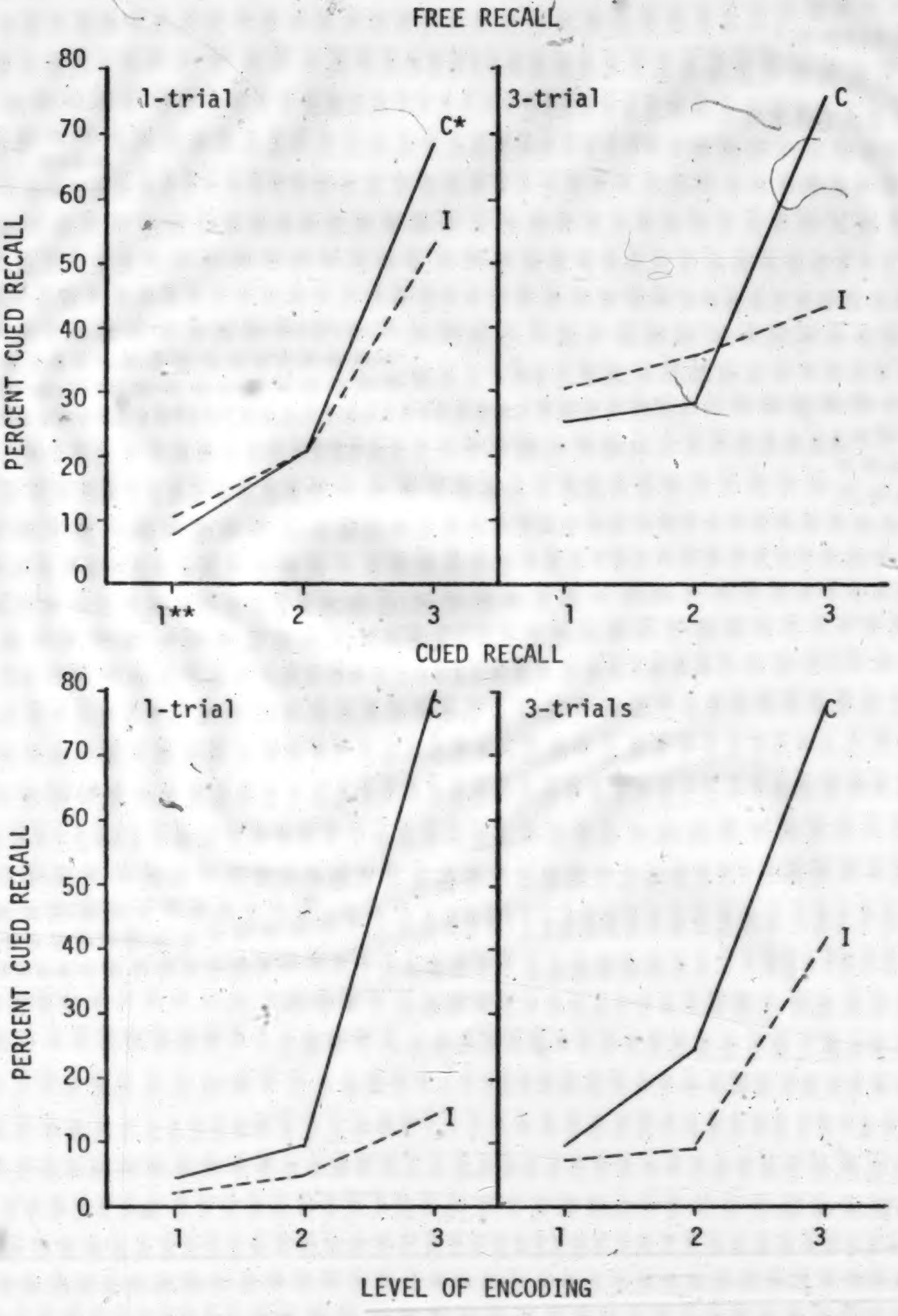
Figure 3. Experiment 3: Cued recall as a function of level of encoding and test condition.



*C = Congruous Encodings; I = Incongruous Encodings

**1 = Structural Encodings; 2 = Phonemic Encodings

Figure 4. Experiment 5: Final cued recall with phonemic cues as a function of level of encoding, congruity, number of trials, and immediate test condition.



*C = Congruous Encodings; I = Incongruous Encodings
**1 = Structural Encodings; 2 = Phonemic Encodings
3 = Semantic Encodings

Figure 5. Experiment 5: Final cued recall with semantic cues as a function of level of encoding, congruity, number of trials, and immediate test condition.

REFERENCES

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123.

Anderson, J. R., & Bower, G. H. Human Associative Memory. Washington, D. C.: Winston, 1973.

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. Journal of Experimental Psychology, 1968, 77, 171-179.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory. Vol. 2, New York: Academic Press, 1968. Pp. 89-195.

Atkinson, R. C., & Shiffrin, R. M. The control of short-term memory. Scientific American, August, 1971, 82-90.

Baddeley, A. D. Retrieval rules and semantic coding in short-term memory. Psychological Bulletin, 1972, 78, 379-385. (b)

Bahrick, H. P. Two-phase model for prompted recall. Psychological Review, 1970, 77, 215-222.

Bartlett, F. C. Remembering. Cambridge: Cambridge University Press, 1932.

Begg, I. Recall of meaningful phrases. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 431-439.

Bellezza, F. S., Richards, D. L., & Geiselman, R. E. Semantic processing and organization in free recall. Memory and Cognition, 1976, 4, 415-421.

Bellezza, F. S., & Walker, R. J. Storage-coding trade-off in short-term store. Journal of Experimental Psychology, 1974, 102, 629-633.

Bernbach, H. A. Rate of presentation in free recall: A problem for two-stage memory theories. _Journal of Experimental Psychology: Human Learning and Memory_, 1975, _1_, 18-22.

Bird, C. P. On the role of processing requirements in short-term memory. _Journal of Experimental Psychology: Human Learning and Memory_, 1976, _2_, 234-243.

Bjork, R. A. Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), _Coding Processes in Human Memory_. Washington, D. C.: Winston, 1972. Pp. 217-235.

Bjork, R. A., & Whitten, W. B. Recency sensitive retrieval processes in long-term free recall. _Cognitive Psychology_, 1974, _6_, 173-189.

Blake, M. Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. _Journal of Verbal Learning and Verbal Behavior_, 1973, _12_, 311-319.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), _The Psychology of Human Learning and Memory: Advances in Research and Theory_. Vol. 1, New York: Academic Press, 1967. Pp. 229-325.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. _Journal of Verbal Learning and Verbal Behavior_, 1969, _8_, 323-343.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. _Cognitive Psychology_, 1971, _2_, 331-350.

Bregman, A. S. Forgetting curves with semantic, phonetic, graphic, and contiguity cues. _Journal of Experimental Psychology_, 1968, _78_, 539-546.

Broadbent, D. E. _Perception and Communication_. New York: Pergamon Press, 1958.

Brodie, D. A. Free recall measures of short-term store: Are rehearsal and order of recall data necessary? Memory and Cognition, 1975, 3, 653-662.

Brown, J. Recall and Recognition. London: Wiley, 1976.

Brown, R., & McNeill, D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 325-337.

Coltheart, V. Recognition errors after incidental learning as a function of different levels of processing. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 437-444.

Conrad, R. Interference or decay over short retention intervals? Journal of Verbal Learning and Verbal Behavior, 1967, 6, 49-54.

Cermak, R. S., & Youtz, C. P. Retention of phonemic and semantic features of words. Memory and Cognition, 1976, 4, 172-175.

Corteen, R. S., & Wood, B. Autonomic responses to shock associated words in an unattended channel. Journal of Experimental Psychology, 1972, 97, 308-313.

Craik, F. I. M. Two components in free recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 996-1004.

Craik, F. I. M. The fate of primary memory items in free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 143-148.

Craik, F. I. M. A "Levels of Analysis" view of memory. In P. Pliner, L. Krames, & T. M. Alloway (Eds.), Communication and Affect: Language and Thought. New York: Academic Press, 1973. Pp. 45-65.

Craik, F. I. M., & Jacoby, L. L. A process view of short-term retention. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), Cognitive Theory. New York: Wiley, 1975. Pp. 173-192.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. _Journal of Verbal Learning and Verbal Behavior_, 1972, _11_, 671-684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. _Journal of Experimental Psychology: General_, 1975, _104_, 268-294.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. _Journal of Verbal Learning and Verbal Behavior_, 1973, _12_, 599-607.

Dornbush, R. L., & Winnick, W. A. Short-term intentional and incidental learning. _Journal of Experimental Psychology_, 1967, _73_, 608-611.

Eagle, M., & Leiter, E. Recall and recognition in intentional and incidental learning. _Journal of Experimental Psychology_, 1964, _68_, 58-63.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal discrimination learning. _Psychological Review_, 1966, _73_, 566-578.

Elias, C. S., & Perfetti, C. A. Encoding task and recognition memory: The importance of semantic coding. _Journal of Experimental Psychology_, 1973, _99_, 151-156.

Estes, W. K., & DaPolito, F. J. Independent variation of information storage and retrieval processes in paired-associate learning. _Journal of Experimental Psychology_, 1967, _75_, 18-26.

Gardiner, J. M. Levels of processing in word recognition and subsequent free recall. _Journal of Experimental Psychology_, 1974, _102_, 101-105.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), _The Psychology of Learning and Motivation: Advances in Research and Theory_. Vol. 5, New York: Academic Press, 1972. Pp. 129-193.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. _Journal of Verbal Learning and Verbal Behavior_, 1966, _5_, 351-360.

Herriot, P. *Attributes of Memory*. London: Methuen, 1974.

Horowitz, L. M., & Prytulak, L. S. Redintegrative memory. *Psychological Review*, 1969, *76*, 519-531.

Horton, K. D. Phonemic similarity, overt rehearsal, and short-term store. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 244-251.

Humphreys, M. S. Relational information and the context effect in recognition memory. *Memory and Cognition*, 1976, *4*, 221-232.

Hyde, T. S., & Jenkins, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, *82*, 472-481.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, *45*, 188-196.

Jacoby, L. L. Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 302-310.

Jacoby, L. L. The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 483-496.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 561-565.

Jacoby, L. L., & Goolkasian, P. Semantic versus acoustic coding: Retention and conditions of organization. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 324-333.

Jenkins, J. J. Remember that old theory of memory? Well forget it! *American Psychologist*, 1974, *29*, 785-795. (a)

Jenkins, J. J. Can we have a theory of meaningful memory? In R. L. Solso (Ed.), Theories in Cognitive Psychology. New York: Wiley, 1974. Pp. 1-20. (b)

Johnston, C. D., & Jenkins, J. J. Two more incidental tasks that differentially affect associative clustering in recall. Journal of Experimental Psychology, 1971, 89, 92-95.

Keele, S. W. The repetition effect: A memory dependent process. Journal of Experimental Psychology, 1969, 80, 243-248.

Kintsch, W. Learning, Memory, and Conceptual Processes. New York: Wiley, 1970.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972. Pp. 247-308. (b)

Kintsch, W. The Representation of Meaning in Memory. Potomac, Maryland: Erlbaum Associates, 1974.

Klein, K., & Saltz, E. Specifying the mechanisms in a levels-of-processing approach to memory. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 671-679.

Kolers, P. A. Remembering operations. Memory and Cognition, 1973, 1, 347-355.

Kuhn, T. S. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1970.

Lackner, J. R., & Garrett, M. I. Resolving ambiguity effects of biasing context in the unattended ear. Cognition, 1973, 1, 359-374.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. Journal of Experimental Psychology, 1970, 85, 225-228.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 1-11.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and Recognition*. London: Wiley, 1976. Pp. 75-102.

Mandler, G. Consciousness: Respectable, useful, and probably necessary. In R. L. Solso (Ed.), *Information Processing and Cognition*. New York: Wiley, 1975. Pp. 229-254.

Mandler, G., & Dean, P. J. Seriation: Development of serial order in free recall. *Journal of Experimental Psychology*, 1969, *81*, 207-215.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, 1968, *75*, 421-441.

Martin, E. Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory*. Toronto: Wiley, 1972. Pp. 59-84.

Maskarinec, A. S., & Brown, S. C. Positive and negative recency effects in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1974, *16*, 328-334.

Mazuryk, G. F. Positive recency in free recall. *Journal of Experimental Psychology*, 1974, *103*, 812-814.

Mazuryk, G. F., & Lockhart, R. S. Negative recency and levels of processing in free recall. *Canadian Journal of Psychology*, 1974, *28*, 114-123.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 1-21.

Meunier, G. F., Kestner, J., Meunier, J. A., & Ritz, D. Overt rehearsal and long-term retention. *Journal of Experimental Psychology*, 1974, *102*, 913-914.

Modigliani, V., & Seamon, J. G. Transfer of information from short- to long-term memory. _Journal of Experimental Psychology_, 1974, _102_, 768-772.

Mondani, M. S., Pellegrino, J. W., & Battig, W. F. Free and cued recall as a function of different levels of word processing. _Journal of Experimental Psychology_, 1973, _101_, 324-329.

Montague, W. E., Adams, J. A., & Kiess, H. O. Forgetting and natural language mediation. _Journal of Experimental Psychology_, 1966, _72_, 829-833.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instructions. _Quarterly Journal of Experimental Psychology_, 1959, _11_, 56-60.

Morris, P. E., & Stevens, R. Linking images and free recall. _Journal of Verbal Learning and Verbal Behavior_, 1974, _13_, 310-315.

Moscovitch, M., & Craik, F. I. M. Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. _Journal of Verbal Learning and Verbal Behavior_, 1976, _15_, 447-458.

Murdock, B. B., Jr. The serial position effect of free recall. _Journal of Experimental Psychology_, 1962, _64_, 482-488.

Murdock, B. B., Jr. Recent developments in short-term memory. _British Journal of Psychology_, 1967, _58_, 421-433.

Murdock, B. B., Jr. _Human Memory: Theory and Data_. New York: Wiley, 1974.

Neisser, U. _Cognitive Psychology_. New York: Appleton-Century-Crofts, 1967.

Neisser, U., & Beller, H. K. Searching through word lists. _British Journal of Psychology_, 1965, _56_, 349-358.

Nelson, D. L., & Borden, R. C. Effect of meaning on processing of phonetic features of words. Journal of Experimental Psychology, 1973, 101, 373-375.

Nelson, D. L., Wheeler, J. W., Jr., Borden, R. C., & Brooks, D. H. Levels of processing and cuing: Sensory versus meaning features. Journal of Experimental Psychology, 1974, 103, 971-977.

Norman, D. A. Toward a theory of memory and attention. Psychological Review, 1968, 75, 522-536.

Norman, D. A. Memory and Attention: An Introduction to Human Information Processing. New York: Wiley, 1969. (a)

Norman, D. A. Memory while shadowing. Quarterly Journal of Experimental Psychology, 1969, 21, 85-93. (b)

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Posner, M. I., & Boies, S. W. Components of attention. Psychological Review, 1971, 78, 391-408.

Posner, M. I., & Konick, A. F. On the role of interference in short-term retention. Journal of Experimental Psychology, 1966, 72, 221-231.

Posner, M. I., & Snyder, C. R. R. Attention and cognitive control. In R. L. Solso (Ed.), Information Processing and Cognition. New York: Wiley, 1975. Pp. 55-85.

Posner, M. I., & Warren, R. E. Traces, concepts, and conscious constructions. In A. W. Melton & E. Martin (Eds.), Coding Processes in Human Memory. Toronto: Wiley, 1972. Pp. 25-43.

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), Categories of Human Learning. New York: Academic Press, 1964. Pp. 146-201.

Prytulak, L. S. Natural language mediation. _Cognitive Psychology_, 1971, _2_, 1-56.

Restle, F. Critique of pure memory. In R. L. Solso (Ed.), _Theories in Cognitive Psychology_. New York: Wiley, 1974. Pp. 203-217.

Roenker, D. L. The role of rehearsal in long-term retention. _Journal of Experimental Psychology_, 1974, _103_, 368-371.

Rowe, E. J., & Paivio, A. Imagery and repetition instructions in verbal discrimination and incidental paired-associate learning. _Journal of Verbal Learning and Verbal Behavior_, 1971, _10_, 668-672.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), _Organization of Memory_. New York: Academic Press, 1972. Pp. 197-245.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. _Journal of Verbal Learning and Verbal Behavior_, 1970, _9_, 99-105.

Runquist, W. N. Interference among memory traces. _Memory and Cognition_, 1975, _3_, 143-159.

Russo, J. E., & Wisher, R. A. Reprocessing as a recognition cue. _Memory and Cognition_, 1976, _4_, 683-689.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. _Perception and Psychophysics_, 1967, _2_, 437-442.

Santa, J. L., & Lamwers, L. L. Encoding specificity: Fact or artifact? _Journal of Verbal Learning and Verbal Behavior_, 1974, _13_, 412-423.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. _Psychological Review_, 1977, _84_, 1-66.

Schulman, A. I. Recognition memory for targets from a scanned word list. British Journal of Psychology, 1971, 62, 335-346.

Schulman, A. I. Memory for words recently classified. Memory and Cognition, 1974, 2, 47-52.

Schulman, A. I. Encoding processes and the memorability of events. In A. Kennedy, & A. Wilkes (Eds.), Studies in Long-Term Memory. Toronto: Wiley, 1975. Pp. 43-56.

Seamon, J. G., & Murray, P. Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 680-687.

Shiffrin, R. M., & Grantham, D. W. Can attention be allocated to sensory modalities? Perception and Psychophysics, 1974, 15, 460-474.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

Shulman, H. G. Encoding and retention of semantic and phonemic information in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 499-508.

Shulman, H. G. Similarity effects in short-term memory. Psychological Bulletin, 1971, 75, 399-415.

Shulman, H. G. Semantic confusion errors in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 221-227.

Smith, E. E., Barresi, J., & Gross, A. E. Imaginal versus verbal coding and the primary-secondary memory distinction. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 597-603.

Smith, M. C., & Groen, M. Evidence for semantic analysis of unattended verbal items. Journal of Experimental Psychology, 1974, 102, 595-603.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Till, R. E., Cormak, D. R., & Prince, P. L. Effects of orienting tasks on sentence comprehension and cued recall. Memory and Cognition, 1977, 5, 59-66.

Till, R. E., Diehl, R. L., & Jenkins, J. J. Effects of semantic and nonsemantic cued orienting tasks on associative clustering in free recall. Memory and Cognition, 1975, 3, 19-23.

Treisman, A. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 1960, 12, 242-248.

Treisman, A. Verbal cues, language, and meaning in attention. American Journal of Psychology, 1964, 77, 206-214.

Treisman, A., & Tuxworth, J. Immediate and delayed recall of sentences after perceptual processing at different levels. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 38-44.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972. Pp. 381-403.

Tulving, E. Ecphoric processes in recall and recognition. In J. Brown (Ed.), Recall and Recognition. London: Wiley, 1976. Pp. 37-73.

Tulving, E., & Colotla, V. A. Free recall of trilingual lists. Cognitive Psychology, 1970, 1, 86-89.

Tulving, E., & Madigan, S. A. Memory and verbal learning. Annual Review of Psychology, 1970, 21, 437-484.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 381-391.

Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, *80*, 352-373.

Tzeng, O. J. L. Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 436-439.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, *76*, 559-573.

Underwood, B. J., & Freund, J. S. Errors in recognition learning and retention. *Journal of Experimental Psychology*, 1968, *78*, 55-63.

Walsh, D. A., & Jenkins, J. J. Effects of orienting tasks on free recall in incidental learning: "Difficulty", "effort", and "process" explanations. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 481-488.

Warrington, E. K., & Ackroyd, C. The effect of orienting tasks on recognition memory. *Memory and Cognition*, 1975, *3*, 140-142.

Watkins, M. J. Concept and measurement of primary memory. *Psychological Bulletin*, 1974, *81*, 695-711. (a)

Watkins, M. J. When is recall spectacularly higher than recognition? *Journal of Experimental Psychology*, 1974, *102*, 161-163. (b)

Watkins, M. J., & Watkins, O. C. Processing of recency items for free recall. *Journal of Experimental Psychology*, 1974, *102*, 488-493.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, *72*, 89-104.

Weingartner, H., Walker, T., Eich, J. E., & Murphy, D. L. Storage and recall of verbal and pictorial information. *Bulletin of the Psychonomic Society*, 1976, *7*, 349-351.

Weist, R. M. The role of rehearsal: Recopy or reconstruct. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 440-450.

Wetzel, C. D. Effect of orienting tasks and cue timing on the free recall of remember- and forget-cued words. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 556-566.

Wickelgren, W. A. The long and the short of memory. Psychological Bulletin, 1973, 80, 425-438.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. Psychological Review, 1970, 77, 1-15.

Wickens, D. D. Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), Coding Processes in Human Memory. Washington, D. C.: Winston, 1972. Pp. 191-215.

Wickens, D. D., Moody, M., & Shearer, P. W. Lack of memory for nonattended items in dichotic listening. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 712-719.

Wicker, F. W., & Bernstein, A. L. Association value and orienting task in incidental and intentional paired-associate learning. Journal of Experimental Psychology, 1969, 81, 308-311.

Wolk, S. The influence of meaningfulness upon intentional and incidental learning of verbal material. Memory and Cognition, 1974, 2, 189-193.

APPENDIX I

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 1: FREE RECALL

SOURCE‡	SS	df	MS	F
I	0.207	2	0.104	<1
S(I)	17.281	42	0.411	
L	39.089	2	19.544	68.03**
R	12.844	1	12.844	44.71**
T	32.941	2	16.470	57.33**
IL	0.881	4	0.220	<1
IR	0.030	2	0.015	<1
LR	18.067	2	9.033	31.44**
IT	1.363	4	0.341	1.19
LT	80.148	4	20.037	69.74**
RT	35.341	2	17.670	61.50**
ILR	0.437	4	0.109	<1
ILT	2.859	8	0.357	1.24
IRT	0.296	4	0.074	<1
LRT	45.481	4	11.370	39.58**
ILRT	1.304	8	0.163	<1
Error	205.157	714	0.287	

** - $p<.01$

‡ - I: Instruction Condition; L: Level of Coding; R: Response Type; T: Type of Encoding Information Recalled

APPENDIX II

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 1: RECOGNITION (HITS)[‡]

SOURCE	SS	df	MS	ERROR TERM	F
I	1.380	2	0.690	S(I)	1.46
S(I)	19.859	42	0.473		
L	30.380	2	15.190	SL(I)	53.45**
R	2.500	1	2.500	SR(I)	11.00**
T	22.706	2	11.353	ST(I)	17.32**
IL	1.079	4	0.270	SL(I)	<1
IR	0.230	2	0.115	SR(I)	<1
LR	1.156	2	0.578	SLR(I)	2.40
IT	1.220	4	0.305	ST(I)	<1
LT	196.731	4	49.183	SLT(I)	47.33**
RT	131.052	2	65.526	SRT(I)	97.36**
ILR	0.837	4	0.209	SLR(I)	<1
ILT	6.009	8	0.751	SLT(I)	<1
IRT	0.963	4	0.241	SRT(I)	<1
LRT	138.547	4	34.637	SLRT(I)	36.21**
ILRT	5.080	8	0.635	SLRT(I)	<1
SL(I)	23.872	84	0.284		
SR(I)	9.548	42	0.227		
ST(I)	55.078	84	0.656		
SLR(I)	20.226	84	0.241		
SLT(I)	174.571	168	1.039		
SRT(I)	56.535	84	0.673		
SLRT(I)	160.719	168	0.957		

** - $p<.01$

‡ - Error terms cannot be pooled due to to heterogeneity of variance: appropriate error terms are noted for each effect.

APPENDIX III

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 2: FREE AND CUED RECALL

SOURCE	SS	df	MS	F
G‡	65011.82	2	32505.91	128.06**
C‡‡	99247.50	1	99247.50	391.01**
GC	6951.613	2	3475.807	13.69**
S(GC)	28936.00	114	253.825	
L	116092.3	2	58046.13	350.90**
GL	11326.88	4	2831.719	17.12**
CL	26988.44	2	13494.22	81.58**
GCL	4909.262	4	1227.315	7.42**
Error	37715.61	228	165.419	

** - $p<.01$
‡G - Study Condition
‡‡C - Test Condition

APPENDIX IV

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 2: FREE RECALL

SOURCE	SS	df	MS	F
G	4913.609	2	2456.805	49.49**
S(G)	2829.641	57	49.643	
L	6380.824	2	3190.412	70.41**
T	21908.60	2	10954.30	241.75**
GL	551.394	4	137.849	3.04*
GT	2606.090	4	651.522	14.38**
LT	22152.21	4	5538.051	122.22**
GLT	1697.172	8	212.146	4.68**
Error	20662.50	456	45.312	

* - $p<.05$
** - $p<.01$

APPENDIX V

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 3: CUED RECALL (ALL GROUPS)

SOURCE	SS	df	MS	F
C	0.608	2	0.304	<1
S(C)	89.524	45	1.989	
E	550.891	4	137.723	161.18**
CE	24.057	8	3.007	3.52**
Error	153.805	180	0.854	

** - p<.01

APPENDIX VI

L

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 4: RECOGNITION

SOURCE	SS	df	MS	F
G	25.176	2	12.588	11.61**
S(G)	55.282	51	1.084	
L	22.594	1	22.594	42.61**
R	15.125	1	15.125	28.53**
D[‡]	0.148	2	0.074	<1
GL	14.373	2	7.186	13.56**
GR	8.676	2	4.338	8.19**
LR	1.680	1	1.680	3.17
GD	0.843	4	0.211	<1
LD	1.123	2	0.561	1.06
RD	0.111	2	0.055	<1
GLR	2.009	2	1.004	1.89
GLD	1.367	4	0.342	<1
GRD	1.769	4	0.442	<1
LRD	0.333	2	0.166	<1
GLRD	0.713	4	0.178	<1
Error	297.464	561	0.530	

** - p<.01

[‡] D - Distractor Type

APPENDIX VII

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 5: IMMEDIATE RECALL ACROSS TRIALS IN THE 3-TRIAL GROUPS

SOURCE	SS	df	MS	F
C	43.512	1	43.512	12.70**
S(C)	130.174	38	3.426	
T	8.808	2	4.404	7.26**
L	242.358	2	121.179	199.87**
R	94.612	1	94.612	156.05**
CT	1.225	2	0.612	1.01
CL	4.075	2	2.037	3.36*
TL	3.133	4	0.783	1.29
CR	3.901	1	3.901	6.43*
TR	1.225	2	0.612	1.01
LR	67.725	2	33.862	55.85**
CTL	1.700	4	0.425	<1
CTR	1.719	2	0.860	1.42
CLR	1.686	2	0.843	1.39
TLR	3.050	4	0.762	1.26
CTLR	0.505	4	0.126	<1
Error	391.647	646	0.606	

* - $p<.05$

** - $p<.01$

APPENDIX VIII

ANALYSIS OF VARIANCE SUMMARY TABLE

EXPERIMENT 5: FINAL CUED RECALL

WITH SEMANTIC CUES

SOURCE	SS	df	MS	F
C	5.002	1	5.002	10.53**
N‡	4.219	1	4.219	8.88**
CN	0.002	1	0.002	<1
S(CN)	36.108	76	0.475	
L	73.017	2	36.508	122.31**
R	8.269	1	8.269	27.70**
CL	0.517	2	0.258	<1
NL	0.350	2	0.175	<1
CR	2.852	1	2.852	9.55**
NR	0.052	1	0.052	<1
LR	13.650	2	6.825	22.86**
CNL	1.517	2	0.758	2.54
CNR	0.252	1	0.252	<1
CLR	0.817	2	0.408	1.37
NLR	0.117	2	0.058	<1
CNLR	2.316	2	1.158	3.88*
Error	113.433	380	0.298	

* - p<.05
** - p<.01
‡N - Number of Trials